Journal of Multivariate Analysis 101 (2010) 1102–1122

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Minimum Hellinger distance estimation in a two-sample semiparametric model

Jingjing Wu [a], Rohana Karunamuni [b,*], Biao Zhang [c]

[a] Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

[b] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

[c] Department of Mathematics, University of Toledo, Toledo, OH 43606-3390, USA

ARTICLE INFO

Article history:
Received 3 December 2008
Available online 22 January 2010

AMS 2000 subject classifications:
Primary 62F10
62E20
secondary 60F05

Keywords:
Asymptotic normality
Hellinger distance
Kernel estimator
Two-sample semiparametric model

ABSTRACT

We investigate the estimation problem of parameters in a two-sample semiparametric model. Specifically, let $X_1, \ldots, X_n$ be a sample from a population with distribution function G and density function g. Independent of the X_i's, let $Z_1, \ldots, Z_m$ be another random sample with distribution function H and density function $h(x) = \exp[\alpha + r(x)\beta]g(x)$, where α and β are unknown parameters of interest and g is an unknown density. This model has wide applications in logistic discriminant analysis, case-control studies, and analysis of receiver operating characteristic curves. Furthermore, it can be considered as a biased sampling model with weight function depending on unknown parameters. In this paper, we construct minimum Hellinger distance estimators of α and β. The proposed estimators are chosen to minimize the Hellinger distance between a semiparametric model and a nonparametric density estimator. Theoretical properties such as the existence, strong consistency and asymptotic normality are investigated. Robustness of proposed estimators is also examined using a Monte Carlo study.

1. Introduction

Semiparametric models have continued to receive increasing attention over the years from both practical and theoretical point of views due in large part to the fact that semiparametric models arise frequently in many areas, primarily in biostatistics and econometrics. The well-known semiparametric models include the Cox proportional hazard model in survival analysis, econometric index models, regression models and errors-in-variables models, among many others. More examples and theory on semiparametric models can be found in the monographs [1,2] and in the review articles [3,4].

In this paper, we consider the following two-sample semiparametric model: Let $X_1, \ldots, X_n$ be a random sample from a population with distribution function G and density function g. Independent of the X_i's, let $Z_1, \ldots, Z_m$ be another random sample from a population with distribution function H and density function h. The two unknown density functions g and h are linked by an "exponential tilt" $\exp[\alpha + r(x)\beta]$. Thus, we have

$$\begin{aligned} &X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} g(x) \\ &Z_1, \ldots, Z_m \overset{\text{i.i.d.}}{\sim} g(x)\exp[\alpha + r(x)\beta], \end{aligned} \tag{1.1}$$

where $r(x) = (r_1(x), \ldots, r_p(x))$ is a $1 \times p$ vector of functions of x, $\beta = (\beta_1, \ldots, \beta_p)^T$ is a $p \times 1$ parameter vector, and α is a normalizing parameter that makes $g(x)\exp[\alpha + r(x)\beta]$ integrate to 1. Various choices of $r(x)$ for some conventional distributions are discussed in [5]. In most applications $r(x) = x$ or $r(x) = (x, x^2)$. Note also that the test of equality of G

* Corresponding author.
E-mail address: R.J.Karunamuni@ualberta.ca (R. Karunamuni).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.01.006

and H can be regarded as a special case of model (1.1) with $\alpha = \beta = 0$. We are interested in the estimation problem of parameters α and β when g is unknown (nuisance parameter).

For $r(x) = x$, model (1.1) encompasses many common distributions, including two exponential distributions with different means and two normal distributions with common variance but different means. Furthermore, model (1.1) with $r(x) = x$ or $r(x) = (x, x^2)$ has wide applications in the logistic discriminant analysis [6,7] and in case-control studies [5,8]. Model (1.1) can also be viewed as a biased sampling model with weight function $\exp[\alpha + r(x)\beta]$ depending on the unknown parameters α and β, see [9]. In [10], a goodness-of-fit test is considered for a logistic regression model based on case-control data by employing the maximum semiparametric likelihood estimator of G to test the validity of model (1.1) with $r(x) = x$. In [11], quantiles of G are estimated under model (1.1).

In this paper, we propose MHD estimation for the two-sample semiparametric model (1.1). In fully parametric models, MHD estimators have been shown to achieve efficiency and have excellent robustness properties such as the resistance to outliers and robustness with respect to model misspecification, see [12,13]. Efficiency combined with excellent robustness properties make MHD estimators appealing in practice. For a comparison between MHD estimators with the MLEs and the balance between robustness and efficiency of estimators, see [14]. Moreover, it has been shown that MLE and MHD estimators are members of a larger class of efficient estimators with various second-order efficiency properties [14]. MHD estimation in fully parametric models have been investigated by various authors, including [12,15–22]. MHD estimators for branching processes and for the mixture complexity in a finite mixture model have been studied in [23–25]. However, MHD estimators for semiparametric models have been studied less. A MHD estimator for finite mixtures of Poisson regression models with the distribution of covariates unknown has been investigated in [26]. Recently, a MHD estimator of the mixture parameter for a nonparametric two-component mixture model has been obtained in [27,28]. Apart from the preceding three articles, there has been very little work reported in the literature on the application of the MHD methodology for semiparametric models. In this paper, we extend the implementation of the MHD approach to the two-sample semiparametric model (1.1). Specifically, we construct minimum Hellinger distance estimators of parameters α and β in model (1.1). The proposed estimators are chosen to minimize the Hellinger distance between a semiparametric model and a nonparametric density estimator. Asymptotic properties such as the existence, strong consistency and asymptotic normality of the proposed MHD estimators of α and β are investigated. Robustness of proposed estimators is also examined using a Monte Carlo study.

This paper is organized as follows. In Section 2, we investigate MHD estimators of the parameters α and β and study their existence and strong consistency. In Section 3, we derive the asymptotic distribution of the proposed estimators. Section 4 contains a simulation study where efficiency and robustness properties of the proposed MHD estimators are studied using a Monte Carlo study. A real data example is given in Section 5. A detailed proof of asymptotic normality of the estimators (Theorem 3.2) is given in Section 6.

2. MHD estimators of regression parameters

Define $\theta = (\alpha, \beta^T)^T$, where α and β are as in (1.1). Then the model (1.1) can be written as

$$\begin{aligned} &X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} g(x) \\ &Z_1, \ldots, Z_m \overset{\text{i.i.d.}}{\sim} h_\theta(x), \end{aligned} \tag{2.1}$$

where $h_\theta(x) = g(x)\exp[(1, r(x))\theta]$, $r(x) = (r_1(x), \ldots, r_p(x))$ is a $1 \times p$ vector of continuous functions of x on $\mathbb{R}$, $\beta = (\beta_1, \ldots, \beta_p)^T$ is a $p \times 1$ parameter vector and α is a normalizing parameter that makes $h_\theta(x)$ integrate to 1. We assume here and in what follows that $\theta \in \Theta$ and Θ is a compact subset of $\mathbb{R}^{p+1}$.

We first define following kernel density estimators of g and h_θ based on the data $X_1, \ldots, X_n$ and $Z_1, \ldots, Z_m$, respectively, of (2.1):

$$g_n(x) = \frac{1}{nb_n}\sum_{i=1}^{n} K_0\left(\frac{x - X_i}{b_n}\right), \tag{2.2}$$

$$h_m(x) = \frac{1}{mb_m}\sum_{j=1}^{m} K_1\left(\frac{x - Z_j}{b_m}\right), \tag{2.3}$$

where K_0 and K_1 are symmetric density functions, bandwidths b_n and b_m are positive constants such that $b_n \to 0$ as $n \to \infty$ and $b_m \to 0$ as $m \to \infty$. We can also employ adaptive kernel density estimators, which use $S_n b_n$ instead of b_n with S_n being a robust scale statistic. Here we use non-adaptive kernel density estimators (2.2) and (2.3) for convenience. The results can be easily extended for adaptive kernel density estimators with some additional conditions on S_n.

Let $\mathcal{H}$ be the set of all densities w.r.t. Lebesgue measure on the real line. For $\phi \in \mathcal{H}$, we define a MHD functional $T_0(\phi)$ as

$$T_0(\phi) = T\left(\{h_\theta\}_{\theta\in\Theta}, \phi\right) = \arg\min_{\theta\in\Theta} \|h_\theta^{1/2} - \phi^{1/2}\|. \tag{2.4}$$

If the family $\{h_\theta\}_{\theta\in\Theta}$ is identifiable, then the functional T_0 is Fisher consistent, i.e., $T_0(h_\theta) = \theta$ for any $\theta \in \Theta$. Since h_m defined by (2.3) is an estimator of h_θ, a MHD estimator of θ will be $T_0(h_m)$. However, this estimator is not available in application since g and hence h_θ in (2.4) are unknown. Naturally, one can use the estimator g_n given by (2.2) in the place of g and then apply the plug-in rule to construct a parametric model, i.e., one replaces h_θ with

$$\widehat{h}_\theta(x) = \exp[(1, r(x))\theta]g_n(x). \tag{2.5}$$

Note that $\widehat{h}_\theta$ is a parametric density function with the unknown parameter being θ. Let $N = n + m$ be the total sample size here and in what follows. Then our proposed MHD estimator of θ is defined as

$$\theta_N = \widehat{T}(h_m) = T\big(\{\widehat{h}_\theta\}_{\theta\in\Theta}, h_m\big) = \arg\min_{\theta\in\Theta} \|\widehat{h}_\theta^{1/2} - h_m^{1/2}\|, \tag{2.6}$$

where h_m and $\widehat{h}_\theta$ are given by (2.3) and (2.5), respectively. That is, θ_N is the minimizer of the Hellinger distance between the parametric density $\widehat{h}_\theta$ and the nonparametric density estimator h_m. This approach is in line with Beran's [12] original mechanism of obtaining MHD estimators. Thus, we would expect θ_N to have good robustness and asymptotic efficiency properties. Since $\widehat{T}(h_m)$ may be multiple valued, the notation $\widehat{T}(h_m)$ is meant to indicate any one of the possible values chosen arbitrarily. Asymptotic properties of θ_N are studied when $n \to \infty$ and $m \to \infty$ as $N \to \infty$.

Note that in (2.6) we are not minimizing the Hellinger distance over a subset of Θ including those θ's which make $\widehat{h}_\theta$ densities, i.e., over $\{\theta \in \Theta : \int \widehat{h}_\theta(x)dx = 1\}$. The reason being that, even for $\theta \in \Theta$ such that $\widehat{h}_\theta$ is not a density, it could make h_θ a density. The true parameter value θ may not make $\widehat{h}_\theta$ a density, but it is not reasonable to exclude θ as an estimate θ_N of itself defined by (2.6). Nevertheless, the definition of θ_N is equivalent to a minimization over a smaller parameter space, as shown in the next lemma. The proofs of lemmas and theorems stated in this section are given in [28,29].

Lemma 2.1. (i) *Suppose that for any $\theta = (\alpha, \beta^T)^T \in \Theta$ there exists $\theta' = (\alpha', \beta^T)^T \in \Theta$ such that $\int \exp[\alpha' + r(x)\beta]g(x)dx = 1$. Let $\Theta_0 = \{\theta \in \Theta : \int \exp[(1, r(x))\theta]g(x)dx \le 1\}$. Then for any $\phi \in \mathcal{H}$,*

$$T_0(\phi) = \arg\min_{\theta\in\Theta} \|h_\theta^{1/2} - \phi^{1/2}\| = \arg\min_{\theta\in\Theta_0} \|h_\theta^{1/2} - \phi^{1/2}\|.$$

(ii) *Suppose that for any $\theta = (\alpha, \beta^T)^T \in \Theta$ there exists $\theta' = (\alpha', \beta^T)^T \in \Theta$ such that $\int \exp[\alpha' + r(x)\beta]g_n(x)dx = 1$. Let $\Theta_n = \{\theta \in \Theta : \int \exp[(1, r(x))\theta]g_n(x)dx \le 1\}$. Then for any $\phi \in \mathcal{H}$,*

$$\widehat{T}(\phi) = \arg\min_{\theta\in\Theta} \|\widehat{h}_\theta^{1/2} - \phi^{1/2}\| = \arg\min_{\theta\in\Theta_n} \|\widehat{h}_\theta^{1/2} - \phi^{1/2}\|,$$

where $\widehat{h}_\theta$ is defined by (2.5).

Remark 2.1. If $\int \exp[(1, r(x))\theta]g(x)dx < \infty$ for any $\theta \in \Theta$ and the parameter space Θ is of the form $\Theta = \mathbb{R} \times \Theta_p$ with $\mathbb{R}$ and Θ_p denoting the parameter spaces for α and β, then the condition in part (i) of Lemma 2.1 holds. Furthermore, if g_n is defined by (2.2) with kernel K_0 compactly supported, then the condition in part (ii) of Lemma 2.1 also holds. Moreover, if $C < \sup_{\beta\in\Theta_p} \int \exp[r(x)\beta]g(x)dx < \infty$ (or $C < \sup_{\beta\in\Theta_p} \int \exp[r(x)\beta]g_n(x)dx < \infty$) for some constant $C > 0$, then the condition in part (i) (or (ii)) of Lemma 2.1 holds with $\Theta = [-M, M] \times \Theta_p$ for some finite positive value M.

We now discuss asymptotic properties of the proposed MHD estimator θ_N. First, some results on the functional $T(\cdot, \cdot)$ (see (2.4)) related to the existence, consistency and asymptotic uniqueness of the MHD estimator of θ are stated. The following condition (D1) and the lemma will be useful to prove above properties.

(D1) There exists an ε-neighborhood $B(\theta, \varepsilon)$ of θ such that $h_t - h_\theta$ is bounded by an integrable function for any $t \in B(\theta, \varepsilon)$.

Lemma 2.2. *If* (D1) *holds for $\theta \in \Theta$, then $d(t) = \|h_t^{1/2} - \phi^{1/2}\|$ is continuous at point $t = \theta$ for any $\phi \in \mathcal{H}$.*

Theorem 2.1. *Suppose that T_0 and $\widehat{T}$ are defined by (2.4) and (2.6), respectively, and* (D1) *holds for all $\theta \in \Theta$. Then*

(i) *For every $\phi \in \mathcal{H}$, there exists $\widehat{T}(\phi) \in \Theta$ satisfying (2.6) with $\widehat{h}_\theta$ and g_n defined by (2.5) and (2.2), respectively, and the kernel K_0 in (2.2) compactly supported. For every $\phi \in \mathcal{H}$, there exists $T_0(\phi) \in \Theta$ satisfying (2.4).*

(ii) *Suppose that $n \to \infty$ and $m \to \infty$ as $N \to \infty$ and $\theta_0 = T_0(\phi)$ is unique. Then $\theta_N = \widehat{T}(\phi_m) \to \theta_0$ as $N \to \infty$ for any density sequences $\{\phi_m\}_{m\in\mathbb{N}}$ and $\{\widehat{h}_\theta\}_{n\in\mathbb{N},\theta\in\Theta}$ such that $\|\phi_m^{1/2} - \phi^{1/2}\| \to 0$ and $\sup_{\theta\in\Theta} \|\widehat{h}_\theta^{1/2} - h_\theta^{1/2}\| \to 0$ as $N \to \infty$.*

(iii) *If $\{h_\theta\}_{\theta\in\Theta}$ is identifiable, then $T_0(h_{\theta_0}) = \theta_0$ uniquely for any $\theta_0 \in \Theta$.*

Remark 2.2. Condition (D1) holds for many families including normal distributions. Suppose that $g(x)$ and $h(x)$ denotes density functions of the normal distributions $N(0, 1)$ and $N(\mu, 1)$, respectively. It is easy to see that $h(x) = h_\theta(x) = \exp[(1, r(x))\theta]g(x)$, where $r(x) = x$ and $\theta = (\alpha, \beta) = (-\frac{\mu^2}{2}, \mu)$. Thus condition (D1) holds for this example.

Remark 2.3. If $\big(1, r(x)\big)$ are linearly independent, then $\{h_\theta\}_{\theta\in\Theta}$ is identifiable. To see this clearly, note that for $h_{\theta_1} = h_{\theta_2}$, we have $(1, r(x))(\theta_1 - \theta_2) = 0$, and so $\theta_1 = \theta_2$ when $(1, r(x))$ are linearly independent. Therefore, $\{h_\theta\}_{\theta\in\Theta}$ is identifiable for any continuous density function g.

With further assumptions on the bandwidths and kernels in (2.2) and (2.3), the consistency of the MHD estimator of θ follows from the continuity of functional T in the Hellinger topology. This result is given next. First, we state a few conditions:

(D2) g and K_0 in (2.1) and (2.2), respectively, have compact supports.

(D3) $\sup_{\theta\in\Theta}\sup_x(1, r(x))\theta < +\infty$.

(D4) g in (2.1) has infinite support, K_0 in (2.2) is a bounded symmetric density with support $[-a_0, a_0]$, $0 < a_0 < \infty$, and there exists a sequence $\{\alpha_n\}$ of positive numbers such that as $n \to \infty$, $\alpha_n \to \infty$ and

$$\sup_{\theta\in\Theta}\int I_{\{|x|>\alpha_n\}}h_\theta(x)\mathrm{d}x \to 0, \tag{2.7}$$

$$b_n^2\sup_{\theta\in\Theta}\int I_{\{|x|>\alpha_n\}}h_\theta(x)\sup_{|t|\le a_0}\frac{|g^{(2)}(x+tb_n)|}{g(x)}\mathrm{d}x \to 0, \tag{2.8}$$

$$n^{-1}b_n^{-1}\sup_{\theta\in\Theta}\int I_{\{|x|\le\alpha_n\}}h_\theta(x)\sup_{|t|\le a_0}\frac{g(x+tb_n)}{g^2(x)}\mathrm{d}x \to 0, \tag{2.9}$$

$$b_n^4\sup_{\theta\in\Theta}\int I_{\{|x|\le\alpha_n\}}h_\theta(x)\sup_{|t|\le a_0}\left[\frac{g^{(2)}(x+tb_n)}{g(x)}\right]^2\mathrm{d}x \to 0, \tag{2.10}$$

where $g^{(k)}$ denotes the kth derivative of g and I_A denotes the indicator function of a set A.

Lemma 2.3. *If* (D4) *holds, then as $n \to \infty$,*

$$\sup_{\theta\in\Theta}\int \exp[(1, r(x))\theta][g_n^{1/2}(x) - g^{1/2}(x)]^2\mathrm{d}x \xrightarrow{P} 0.$$

Theorem 2.2. *Let $n \to \infty$ and $m \to \infty$ as $N \to \infty$. Suppose that $\big(1, r(x)\big)$ are linearly independent,* (D1) *holds for any $\theta \in \Theta$, and the bandwidths b_n and b_m in (2.2) and (2.3), respectively, satisfy $b_n, b_m \to 0$ and $nb_n, mb_m \to \infty$ as $N \to \infty$. Further, suppose that either* (D2), (D3) or (D4) *holds. Then $\|h_m^{1/2} - h_\theta^{1/2}\| \xrightarrow{P} 0$ and $\sup_{\theta\in\Theta}\|\widehat{h}_\theta^{1/2} - h_\theta^{1/2}\| \xrightarrow{P} 0$ as $N \to \infty$. Furthermore, $\theta_N \xrightarrow{P} \theta$ as $N \to \infty$, where θ_N is defined by (2.6) with g_n, h_m and $\widehat{h}_\theta$ given by (2.2), (2.3) and (2.5) respectively.*

Remark 2.4. Condition (D3) is satisfied when g and h_θ are two normal density functions with different standard deviations. For example, assume that $g(x)$ and $h(x)$ denote density functions of $N(0, 1)$ and $N(\mu, \sigma)$, respectively, where $0 < \sigma < 1$. It is easy to see that $h(x) = h_\theta(x) = \exp[(1, r(x))\theta]g(x)$, where $r_1(x) = x$, $r_2(x) = x^2$ and $\theta = (\theta_0, \theta_1, \theta_2) = (-\frac{\mu^2}{2\sigma^2} - \log\sigma, \frac{\mu}{\sigma^2}, \frac{1}{2} - \frac{1}{2\sigma^2})$. If the parameter space Θ is such that its projection onto the third argument is to the left of zero, then clearly condition (D3) holds.

Remark 2.5. Condition (D4) holds for many families and one such example is stated in Remark 2.2, i.e., g and h are two normal density functions with the same standard deviation. Without loss of generality, we suppose the compact parameter space $\Theta = [\underline{\alpha}, \bar{\alpha}] \times [\underline{\beta}, \bar{\beta}]$ for some finite numbers $\bar{\alpha}$, $\underline{\alpha}$, $\bar{\beta}$ and $\underline{\beta}$. Then it is easy to show that (2.7)–(2.10) hold for some α_n, a log function of n, and any bandwidth b_n such that $b_n \to 0$ and $nb_n \to \infty$ as $n \to \infty$.

3. Asymptotic normality

In this section, we obtain the asymptotic distribution of the proposed MHD estimator θ_N. We first state following conditions:

(D5) There exists $B(\theta, \epsilon)$, an ϵ-neighborhood of θ for some $\epsilon > 0$, such that for $s = 1, 2$ and $i, j, k = 0, 1, \ldots, p$,

$$\sup_{t\in\Theta\cap B(\theta,\epsilon)}\sup_x\exp\left[\frac{1}{s}(1, r(x))t\right]|r_i(x)r_j(x)r_k(x)| < \infty,$$

where $r_0(x) = 1$.

(D6) There exists $B(\theta, \epsilon)$, an ϵ-neighborhood of θ for some $\epsilon > 0$, such that for $s = 1, 2$, $i, j, k = 0, 1, \ldots, p$, and $r_0(x) = 1$

$$\int |r_i(x)r_j(x)|^2\exp[(1, r(x))\theta]h_\theta(x)\mathrm{d}x < \infty, \tag{3.1}$$

$$\int |r_i(x)r_j(x)r_k(x)|^s\sup_{t\in\Theta\cap B(\theta,\epsilon)}\exp[(1, r(x))t]\sup_{|t|\le a_0}g(x+tb_n)\mathrm{d}x = O(1), \quad \text{as } n \to \infty, \tag{3.2}$$

$$\int |r_i(x)r_j(x)|^2\exp[2(1, r(x))\theta]\sup_{|t|\le a_0}g(x+tb_n)\mathrm{d}x = O(1), \quad \text{as } n \to \infty. \tag{3.3}$$

Under condition (D2), (D5) or (D6), we derive an expression for the bias term $\theta_N - \theta$, which is presented in the next theorem. We denote $I(\theta) = \int(1, r(x))^T(1, r(x))h_\theta(x)\mathrm{d}x$ and assume that $I(\theta)$ is finite and nonsingular.

Theorem 3.1. *Suppose that $\theta \in int(\Theta)$, K_0 in (2.2) has compact support, and assumptions in Theorem 2.2 hold. Further suppose that either* (D2), (D5) *or* (D6) *holds. Then, it follows that*

$$\theta_N - \theta = \left[I^{-1}(\theta) + \mu_N\right] \times 2 \int \left\{\exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x) h_m^{1/2}(x) - \exp[(1, r(x))\theta] g_n(x)\right\} (1, r(x))^T \mathrm{d}x \tag{3.4}$$

where θ_N is defined by (2.6) and μ_N is a $(p+1) \times (p+1)$ matrix with elements tending to zero in probability as $N \to \infty$.

Remark 3.1. An example in which condition (D5) holds is stated in Remark 2.4. In that example $\theta = (\theta_0, \theta_1, \theta_2)$ with $\theta_2 < 0$. Therefore, one can easily prove that condition (D5) is satisfied. It is also clear that $I(\theta)$ is finite in this case. Condition (D6) is satisfied for the example stated in Remark 2.2, i.e., two normal distributions with the same standard deviation.

In order to state the next theorem, which establishes the asymptotic distribution of the proposed MHD estimator θ_N of θ, a few more conditions are required:

Let $\{\alpha_N\}$ be a sequence of positive numbers such that $\alpha_N \to \infty$ as $N \to \infty$, and

(C0) g has infinite support $(-\infty, \infty)$.

(C1) The second derivatives of g and h_θ exist.

(C2) $n/N \to \rho \in (0, 1)$ as $N \to \infty$.

(C3) K_0 and K_1 in (2.2) and (2.3), respectively, are bounded symmetric densities with supports $[-a_0, a_0]$ and $[-a_1, a_1]$, respectively, where $0 < a_0, a_1 < \infty$.

(C4) $I(\theta) = \int (1, r(x))^T (1, r(x)) h_\theta(x) \mathrm{d}x$ and $J(\theta) = \int (1, r(x))^T (1, r(x)) \exp[(1, r(x))\theta] h_\theta(x) \mathrm{d}x$ are finite.

(C5) The second derivative of g exists and satisfies for $i = 0, 1, \ldots, p$,

$$b_n^2 \int \varepsilon_{Ni}^2(x) h_\theta(x) \sup_{|t| \le a_0} \frac{|g^{(2)}(x + tb_n)|}{g(x)} \mathrm{d}x = O(1) \quad \text{as } N \to \infty,$$

where $\varepsilon_N(x) = (1, r(x))^T I_{\{|x| > \alpha_N\}} = (\varepsilon_{N0}(x), \varepsilon_{N1}(x), \ldots, \varepsilon_{Np}(x))^T$ and $g^{(k)}$ denotes the kth derivative of g.

(C5′) The second derivative of g exists and satisfies

$$N^{1/2} b_n^2 \int |\varepsilon_N(x)| h_\theta(x) \sup_{|t| \le a_0} \frac{|g^{(2)}(x + tb_n)|}{g(x)} \mathrm{d}x = o(1) \quad \text{as } N \to \infty.$$

(C6)

$$N \cdot P(|Z_1| > \alpha_N - a_1 b_m) \to 0 \quad \text{as } N \to \infty,$$
$$N \cdot P(|X_1| > \alpha_N - a_0 b_n) \to 0 \quad \text{as } N \to \infty.$$

(C7)

$$N^{-1/2} b_m^{-1} \int |\delta_N(x)| h_\theta(x) \sup_{|t| \le a_1} \frac{h_\theta(x + tb_m)}{h_\theta^2(x)} \mathrm{d}x \to 0 \quad \text{as } N \to \infty,$$

$$N^{1/2} b_m^4 \int |\delta_N(x)| h_\theta(x) \sup_{|t| \le a_1} \left[\frac{h_\theta^{(2)}(x + tb_m)}{h_\theta(x)}\right]^2 \mathrm{d}x \to 0 \quad \text{as } N \to \infty,$$

$$N^{-1/2} b_n^{-1} \int |\delta_N(x)| h_\theta(x) \sup_{|t| \le a_0} \frac{g(x + tb_n)}{g^2(x)} \mathrm{d}x \to 0 \quad \text{as } N \to \infty,$$

$$N^{1/2} b_n^4 \int |\delta_N(x)| h_\theta(x) \sup_{|t| \le a_0} \left[\frac{g^{(2)}(x + tb_n)}{g(x)}\right]^2 \mathrm{d}x \to 0 \quad \text{as } N \to \infty,$$

where $\delta_N(x) = (1, r(x))^T I_{\{|x| \le \alpha_N\}} = (\delta_{N0}(x), \delta_{N1}(x), \ldots, \delta_{Np}(x))^T$.

(C8)

$$N^{1/2} b_m^2 \int |\delta_N(x)| h_\theta(x) \sup_{|t| \le a_1} \frac{|h_\theta^{(2)}(x + tb_m)|}{h_\theta(x)} \mathrm{d}x \to 0 \quad \text{as } N \to \infty,$$

$$N^{1/2} b_n^2 \int |\delta_N(x)| h_\theta(x) \sup_{|t| \le a_0} \frac{|g^{(2)}(x + tb_n)|}{g(x)} \mathrm{d}x \to 0 \quad \text{as } N \to \infty.$$

(C9)

$$\sup_{|x| \le \alpha_N} \sup_{|t| \le a_1} \frac{h_\theta(x + tb_m)}{h_\theta(x)} = O(1) \quad \text{as } N \to \infty,$$

$$\sup_{|x| \le \alpha_N} \sup_{|t| \le a_0} \frac{g(x + tb_n)}{g(x)} = O(1) \quad \text{as } N \to \infty.$$

(C10) $r(x)$ is differentiable and satisfies for $i = 0, 1, \ldots, p$,

$$b_m^2 \int I_{\{|x|\le\alpha_N\}} h_\theta(x) \sup_{|t|\le a_1} \left(r_i^{(1)}(x+tb_m)\right)^2 \mathrm{d}x \to 0 \quad \text{as } N \to \infty,$$

$$b_n^2 \int I_{\{|x|\le\alpha_N\}} g(x) \sup_{|t|\le a_0} \left[\frac{\partial r_i(y)\exp[(1, r(y))\theta]}{\partial y}\Big|_{y=x+tb_n}\right]^2 \mathrm{d}x \to 0 \quad \text{as } N \to \infty.$$

(C11)

$$N^{-1/2} b_m^{-1} \int |\delta_N(x)| \exp\left[\frac{1}{2}(1, r(x))\theta\right] \mathrm{d}x \to 0 \quad \text{as } N \to \infty,$$

$$N^{1/2} b_m^{4} \int |\delta_N(x)| \exp\left[\frac{1}{2}(1, r(x))\theta\right] \mathrm{d}x \to 0 \quad \text{as } N \to \infty.$$

Theorem 3.2. *Suppose that θ_N defined by (2.6) satisfies (3.4). Further suppose that conditions* (C0)–(C10) *and* (C5′) *hold. Then the asymptotic distribution of $N^{1/2}(\theta_N - \theta)$ is normal with mean 0 and variance Σ, where Σ is defined by*

$$\Sigma = I^{-1}(\theta)\left[\frac{1}{\rho}\Sigma_0 + \frac{1}{1-\rho}\Sigma_1\right] I^{-1}(\theta) \tag{3.5}$$

with

$$\Sigma_0 = \int (1, r(x))^T (1, r(x)) \exp[(1, r(x))\theta] h_\theta(x)\mathrm{d}x - \int (1, r(x))^T h_\theta(x)\mathrm{d}x \int (1, r(x)) h_\theta(x)\mathrm{d}x \tag{3.6}$$

and

$$\Sigma_1 = \int (1, r(x))^T (1, r(x)) h_\theta(x)\mathrm{d}x - \int (1, r(x))^T h_\theta(x)\mathrm{d}x \int (1, r(x)) h_\theta(x)\mathrm{d}x. \tag{3.7}$$

The proof of Theorem 3.2 is given in Section 6 below. In the next few remarks, we discuss the conditions (C0)–(C11) and examine their validity in some examples.

Remark 3.2. Conditions (C0), (C1) and (C4) are typical assumptions on the distributions, and (C3) is also a typical condition on kernels. Conditions (C5)–(C11) and (C5′) basically require that there exists a sequence α_N that controls the tails of the underlying densities. For example, one can easily choose the sequence α_N such that

$$G(\alpha_N) = 1 + o(N^{-1}), \qquad G(-\alpha_N) = o(N^{-1}),$$
$$H_\theta(\alpha_N) = 1 + o(N^{-1}), \qquad H_\theta(-\alpha_N) = o(N^{-1}),$$

then condition (C6) holds, where G and H_θ are the cumulative distribution functions of densities g and h_θ, respectively. Conditions (C0)–(C11) and (C5′) hold for families such as normal distributions.

Condition (C0) requires that g possesses a support $(-\infty, \infty)$. However the results in Theorem 3.2 can be easily applied to any type of infinite support. For example, the exponential distributions have an infinite support $[0, \infty)$. Consider two exponential distributions with densities $\bar{g}(x) = \mathrm{e}^{-x}$ and $\bar{h}(x) = \lambda \mathrm{e}^{-\lambda x}$ with $\lambda > 1$ as an example. Then it is easy to see that $\log X_i$ and $\log Z_i$ are distributed as $g(x) = \exp\{x - \mathrm{e}^x\}$ and $h(x) = \lambda \exp\{x - \lambda \mathrm{e}^x\}$, respectively. Both g and h have the support $(-\infty, \infty)$, and h can be represented as $h_\theta(x) = \exp\{(1, r(x))\theta\} g(x)$ with $r(x) = \mathrm{e}^x$ and $\theta = (\log\lambda, 1-\lambda)^T$. Now if we choose $b_n = O(N^{-r})$ with $1/8 < r < 1/4$ and $\alpha_N = o(\log N)$, then conditions (C0)–(C11) and (C5′) are satisfied, see Remark 3.3 below for detailed calculation of a simpler example.

Remark 3.3. Consider the example stated in Remark 2.2 again. It is easy to see that conditions (C0), (C1) and (C4) hold. We can easily choose bandwidths b_n, b_m and kernels K_0, K_1 satisfying conditions (C2) and (C3). Since for $k = 0, 1, 2$,

$$\int |x|^k h_\theta(x) \sup_{|t|\le a_0} \frac{|g^{(2)}(x+tb_n)|}{g(x)} \mathrm{d}x = O(1) \quad \text{as } N \to \infty,$$

conditions (C5) and (C5′) hold if $Nb_n^4 = O(1)$ as $N \to \infty$. Note that as $N \to \infty$,

$$N \int_{\alpha_N}^{\infty} \exp[-x^2/2]\mathrm{d}x \le N \int_{\alpha_N}^{\infty} x \exp[-x^2/2]\mathrm{d}x = N \exp[-\alpha_N^2/2].$$

Thus, if $N \exp[-\alpha_N^2/2] \to 0$ as $N \to \infty$, then condition (C6) holds. Since for $i = 0, 1$ and $j = 1, 2$,

$$\int |x|^i h_\theta(x) \sup_{|t|\le a_1} \left|\frac{h_\theta^{(2)}(x+tb_m)}{h_\theta(x)}\right|^j \mathrm{d}x = O(1) \quad \text{as } N \to \infty$$

and

$$\int |x|^i h_\theta(x) \sup_{|t|\le a_0} \left|\frac{g^{(2)}(x+tb_n)}{g(x)}\right|^j \mathrm{d}x = O(1) \quad \text{as } N \to \infty,$$

(C8) and the second and fourth expressions in (C7) hold if $Nb_n^4 \to 0$ as $N \to \infty$. If $b_n\alpha_N \to 0$ and $N^{-1/2}b_n^{-1}\alpha_N^2 \to 0$ as $N \to \infty$, then for $i = 0, 1$ as $N \to \infty$,

$$\begin{aligned}
N^{-1/2}b_m^{-1}\int_{-\alpha_N}^{\alpha_N} |x|^i h_\theta(x) \sup_{|t|\le a_1}\frac{h_\theta(x+tb_m)}{h_\theta^2(x)}\mathrm{d}x &= N^{-1/2}b_m^{-1}\int_{-\alpha_N}^{\alpha_N}|x|^i \sup_{|\epsilon|\le a_1 b_m}\exp\left[-\epsilon x+\epsilon\mu-\frac{\epsilon^2}{2}\right]\mathrm{d}x\\
&\le 2\exp[a_1 b_m|\mu|]\cdot N^{-1/2}b_m^{-1}\int_0^{\alpha_N}|x|^i\exp[a_1 b_m x]\mathrm{d}x\\
&\le \frac{2}{a_1}\exp[a_1 b_m|\mu|]\cdot N^{-1/2}b_m^{-2}\alpha_N^i\big(\exp[a_1 b_m\alpha_N]-1\big)\\
&= O\big(N^{-1/2}b_m^{-1}\alpha_N^{i+1}\big)\\
&\to 0,
\end{aligned}$$

and therefore the first expression in (C7) holds. Similarly, for $i = 0, 1$ as $N \to \infty$,

$$\begin{aligned}
N^{-1/2}b_n^{-1}\int_{-\alpha_N}^{\alpha_N}|x|^i h_\theta(x)\sup_{|t|\le a_0}\frac{g(x+tb_n)}{g^2(x)}\mathrm{d}x &= N^{-1/2}b_n^{-1}\int_{-\alpha_N}^{\alpha_N}|x|^i\exp\left[\mu x-\frac{\mu^2}{2}\right]\sup_{|\epsilon|\le a_0 b_n}\exp\left[-\epsilon x-\frac{\epsilon^2}{2}\right]\mathrm{d}x\\
&\le N^{-1/2}b_n^{-1}\alpha_N^i\int_0^{\alpha_N}\exp[(\mu+a_0b_n)x]\mathrm{d}x\\
&\quad + N^{-1/2}b_n^{-1}\alpha_N^i\int_{-\alpha_N}^0\exp[(\mu-a_0b_n)x]\mathrm{d}x\\
&= N^{-1/2}b_n^{-1}\alpha_N^i(\mu+a_0b_n)^{-1}\big(\exp[(\mu+a_0b_n)\alpha_N]-1\big)\\
&\quad + N^{-1/2}b_n^{-1}\alpha_N^i(\mu-a_0b_n)^{-1}\big(1-\exp[-(\mu-a_0b_n)\alpha_N]\big)\\
&= \begin{cases} O\big(N^{-1/2}b_n^{-1}\alpha_N^i\exp[|\mu|\alpha_N]\big) & \text{if } \mu\ne 0,\\ O\big(N^{-1/2}b_n^{-1}\alpha_N^{i+1}\big) & \text{if } \mu = 0.\end{cases}
\end{aligned}$$

Therefore, if $N^{-1/2}b_n^{-1}\alpha_N\exp[|\mu|\alpha_N] \to 0$ as $N \to \infty$, then the third expression in (C7) holds. If $b_n\alpha_N = O(1)$ as $N \to \infty$, then (C9) holds. It is easy to check that (C10) is satisfied. Note that as $N \to \infty$,

$$\int_{-\alpha_N}^{\alpha_N}\exp\left[\frac{1}{2}(1, r(x))\theta\right]\mathrm{d}x = \begin{cases} O\big(\exp[|\mu|\alpha_N/2]\big) & \text{if } \mu\ne 0,\\ O\big(\alpha_N\big) & \text{if } \mu = 0,\end{cases}$$

and

$$\int_{-\alpha_N}^{\alpha_N}|x|\exp\left[\frac{1}{2}(1, r(x))\theta\right]\mathrm{d}x = \begin{cases} O\big(\alpha_N\exp[|\mu|\alpha_N/2]\big) & \text{if } \mu\ne 0,\\ O\big(\alpha_N^2\big) & \text{if } \mu = 0.\end{cases}$$

So if $N^{-1}b_m^{-2}\alpha_N^2\exp[|\mu|\alpha_N] \to 0$ and $Nb_m^4\alpha_N^2\exp[|\mu|\alpha_N] \to 0$ as $N \to \infty$, then (C11) hold. In summary, if we choose

$$b_n = O\big(N^{-r}\big), \quad 1/4 < r < 1/2$$

and

$$\alpha_N = O\big((\log N)^q\big), \quad 1/2 < q < 1,$$

then conditions (C0)–(C10) and (C5′) are satisfied. Also by Remarks 2.2, 2.5 and 3.2, (3.4) holds. As a result, (3.5) holds by Theorem 3.2.

Remark 3.4. We discuss the same example studied in Remark 3.3 here again. Simple calculation yields that the asymptotic variance for our proposed estimator θ_N of θ is

$$\Sigma = \frac{1}{\rho}\begin{bmatrix}\mu^4\exp[\mu^2]-\mu^2\exp[\mu^2]+\exp[\mu^2]-1 & -\mu^3\exp[\mu^2]\\ -\mu^3\exp[\mu^2] & \mu^2\exp[\mu^2]+\exp[\mu^2]\end{bmatrix} + \frac{1}{1-\rho}\begin{bmatrix}\mu^2 & -\mu\\ -\mu & 1\end{bmatrix}. \tag{3.8}$$

Zhang [11] estimated $\theta = (\alpha, \beta)$ using the maximum semiparametric likelihood method for model (1.1). He derived the asymptotic variance, say $\overline{\Sigma}$, of his estimator of θ. It is somewhat difficult to give an explicit expression for the asymptotic variance $\overline{\Sigma}$ in this example. Thus, here we compare asymptotic variances in the simplest case when $\mu = 0$. If $\mu = 0$ then the asymptotic variance of our proposed estimator θ_N is

$$\Sigma = \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{1}{\rho(1-\rho)} \end{bmatrix},$$

which is exactly the same as $\overline{\Sigma}$. More detailed comparison of Σ with $\overline{\Sigma}$ is given in Section 4 below.

Remark 3.5. A MHD estimator could be defined similarly for multivariate observations $X_1, \ldots, X_n, Z_1, \ldots, Z_m \in \mathbb{R}^d$ as well. In the multivariate case, one needs to use the multivariate version of the kernel density estimators defined in (2.2) and (2.3):

$$g_n(x) = \frac{1}{nb_n^d} \sum_{i=1}^{n} \prod_{j=1}^{d} K_0\left(\frac{x_j - X_{ij}}{b_n}\right),$$

$$h_m(x) = \frac{1}{mb_m^d} \sum_{i=1}^{m} \prod_{j=1}^{d} K_1\left(\frac{x_j - Z_{ij}}{b_m}\right).$$

For simplicity here we used common bandwidths for each of the d components. To obtain an asymptotic result as in Theorem 3.2, one needs some higher-order convergence rates involving α_N, a sequence of positive numbers in $\mathbb{R}^d$. Conditions (C5), (C5′), (C7), (C8) and (C10) hold for each partial derivatives of the underlying distributions, g or h_θ. For example, condition (C5) now becomes, for $j, k = 1, 2, \ldots, d$,

$$b_n^2 \int \varepsilon_{Ni}^2(x) h_\theta(x) \sup_{|t_1|,|t_2| \le a_0} \frac{|g^{(jk)}(x + t_1 e_j + t_2 e_k)|}{g(x)} \mathrm{d}x = O(1) \quad \text{as } N \to \infty,$$

where $g^{(jk)}(x) = \frac{\partial^2 g(x)}{\partial x_j \partial x_k}$ and e_j is a $d \times 1$ vector with 1 for the jth component and 0 for others. Furthermore, conditions of (C7) hold with b_m^{-1} and b_n^{-1} replaced by b_m^{-d} and b_n^{-d}, respectively.

Remark 3.6. For the model (1.1), to test whether the two samples come from the same population is equivalent to testing hypotheses $H_0 : \theta = 0$ vs $H_1 : |\theta| > 0$, or simply test $H_0 : \beta = 0$ vs $H_1 : |\beta| > 0$ since α is a function of β. A test statistic could be easily constructed using the asymptotic result in Theorem 3.2 with an appropriate estimate of the asymptotic variance. More specifically, one may use $N^{1/2}\theta_N \widehat{\Sigma}^{-1/2}$ as a test statistic which is approximately normally distributed for large N, where $\widehat{\Sigma}$ is an estimate of Σ defined by (3.5) with g, h_θ, ρ and θ replaced by g_n, h_m, n/N and θ_N respectively. A corresponding confidence interval for θ would then be $(\theta_N - z^* N^{-1/2} \widehat{\Sigma}^{1/2}, \theta_N + z^* N^{-1/2} \widehat{\Sigma}^{1/2})$ with z^* being the z-value corresponding to the confidence level. For the example given in Section 5 below,

$$\Sigma = \begin{bmatrix} 107.779 & -2.246 \\ -2.246 & 0.047 \end{bmatrix},$$

and 95% individual confidence intervals for β and α are $(0.05, 0.13)$ and $(-6.67, -2.61)$, respectively. A more detailed discussion on hypothesis testing under model (2.1) will be presented in a separate paper that is under preparation.

It is well-known that Wilcoxon's two-sample test can be used to compare two populations. However, the preceding test is not capable of detecting differences in the two underlying distributions g and h_θ completely. For instance, when g and h_θ differ only in variation but with same means, Wilcoxon's test will conclude that the two populations are the same. On the other hand, the test based on the MHD estimator proposed above can detect any difference in g and h_θ. Specifically, a polynomial function can be employed to approximate the logarithm of the ratio h_θ/g, and then by letting $r(x)$ be a polynomial function, one can model any difference in g and h_θ.

Remark 3.7. For numerical calculation of the proposed MHD estimator, one may use Newton–Raphson iteration method. For an initial value of θ, one can use $(1, r(z))\theta$ to fit the points $\log h_m(Z_j)/g_n(Z_j), j = 1, \ldots, m$. The preceding method can also be used to obtain a rough idea about the domain Θ of the parameter θ. Alternatively, maximum semiparametric likelihood estimator may be implemented as an initial value of θ. If an empirical parameter space Θ is available, then the minimization will be much easier and one has to simply employ the traversal method for small parameter space Θ. To the best of authors' knowledge, there are no free codes available to test the MHD method. For simplicity, we used $\Theta = [-10, 10]^{p+1}$ in our simulation and $C/C++$ programming.

4. Simulation studies

In this section, we report the results of a Monte Carlo study. In particular, we plan to demonstrate numerically that the proposed MHD estimator θ_N defined in (2.6) has good robustness and efficiency properties.

Table 1

The asymptotic variance matrixes Σ and $\overline{\Sigma}$ of θ_N and $\widetilde{\theta}$ defined in (2.6) and [11], respectively, when g and h are the densities of $N(0, 1)$ and $N(\mu, 1)$, respectively.

ρ	$\mu = 0.1$		$\mu = 0.5$		$\mu = 1$	
	Σ	$\overline{\Sigma}$	Σ	$\overline{\Sigma}$	Σ	$\overline{\Sigma}$
1/6	$\begin{bmatrix} 0.01 & -0.13 \\ -0.13 & 7.32 \end{bmatrix}$	$\begin{bmatrix} 0.01 & -0.12 \\ -0.12 & 7.22 \end{bmatrix}$	$\begin{bmatrix} 0.56 & -1.56 \\ -1.56 & 10.83 \end{bmatrix}$	$\begin{bmatrix} 0.33 & -0.73 \\ -0.73 & 7.74 \end{bmatrix}$	$\begin{bmatrix} 11.51 & -17.51 \\ -17.51 & 33.82 \end{bmatrix}$	$\begin{bmatrix} 1.74 & -2.33 \\ -2.33 & 9.70 \end{bmatrix}$
2/6	$\begin{bmatrix} 0.02 & -0.15 \\ -0.15 & 4.56 \end{bmatrix}$	$\begin{bmatrix} 0.02 & -0.15 \\ -0.15 & 4.52 \end{bmatrix}$	$\begin{bmatrix} 0.50 & -1.23 \\ -1.23 & 6.32 \end{bmatrix}$	$\begin{bmatrix} 0.41 & -0.88 \\ -0.88 & 5.01 \end{bmatrix}$	$\begin{bmatrix} 6.65 & -9.65 \\ -9.65 & 17.81 \end{bmatrix}$	$\begin{bmatrix} 2.02 & -2.54 \\ -2.54 & 6.67 \end{bmatrix}$
3/6	$\begin{bmatrix} 0.02 & -0.20 \\ -0.20 & 4.04 \end{bmatrix}$	$\begin{bmatrix} 0.02 & -0.20 \\ -0.20 & 4.02 \end{bmatrix}$	$\begin{bmatrix} 0.59 & -1.32 \\ -1.32 & 5.21 \end{bmatrix}$	$\begin{bmatrix} 0.53 & -1.13 \\ -1.13 & 4.50 \end{bmatrix}$	$\begin{bmatrix} 5.44 & -7.44 \\ -7.44 & 12.87 \end{bmatrix}$	$\begin{bmatrix} 2.55 & -3.05 \\ -3.05 & 6.09 \end{bmatrix}$
4/6	$\begin{bmatrix} 0.03 & -0.30 \\ -0.30 & 4.53 \end{bmatrix}$	$\begin{bmatrix} 0.03 & -0.30 \\ -0.30 & 4.52 \end{bmatrix}$	$\begin{bmatrix} 0.81 & -1.74 \\ -1.74 & 5.41 \end{bmatrix}$	$\begin{bmatrix} 0.78 & -1.63 \\ -1.63 & 5.01 \end{bmatrix}$	$\begin{bmatrix} 5.58 & -7.08 \\ -7.08 & 11.15 \end{bmatrix}$	$\begin{bmatrix} 3.62 & -4.13 \\ -4.13 & 6.67 \end{bmatrix}$
5/6	$\begin{bmatrix} 0.06 & -0.60 \\ -0.60 & 7.22 \end{bmatrix}$	$\begin{bmatrix} 0.06 & -0.60 \\ -0.60 & 7.22 \end{bmatrix}$	$\begin{bmatrix} 1.55 & -3.19 \\ -3.19 & 7.93 \end{bmatrix}$	$\begin{bmatrix} 1.54 & -3.14 \\ -3.14 & 7.74 \end{bmatrix}$	$\begin{bmatrix} 8.06 & -9.26 \\ -9.26 & 12.52 \end{bmatrix}$	$\begin{bmatrix} 6.78 & -7.37 \\ -7.37 & 9.70 \end{bmatrix}$

In this simulation study, we considered the example discussed in Remark 2.2. We assumed that $g(x)$ and $h(x)$ denote density functions of the normal distributions $N(0, 1)$ and $N(\mu, 1)$, respectively. Thus $h(x) = h_\theta(x) = \exp[(1, r(x))\theta]g(x)$, where $r(x) = x$ and $\theta = (\alpha, \beta) = (-\frac{\mu^2}{2}, \mu)$. For different μ and ρ values, Table 1 compares Σ defined in (3.5) with the asymptotic variance matrix $\overline{\Sigma}$ of the maximum semiparametric likelihood estimator $\widetilde{\theta} = (\widetilde{\alpha}, \widetilde{\beta})$ of [11]. From Table 1, it is easily seen that smaller values of μ result in smaller variance of the estimator $\theta_N = (\widehat{\alpha}, \widehat{\beta})$. The correlations are all negative since $\alpha = -\frac{\beta^2}{2}$. For $\mu = 0$, the asymptotic variance of θ_N is exactly the same as that of $\widetilde{\theta}$ (3.8), and for $\mu = 0.1$, the asymptotic variance of θ_N is almost the same as that of $\widetilde{\theta}$ for all different values of ρ. On the other hand, for large values of μ, the asymptotic variance of θ_N is larger than that of $\widetilde{\theta}$. In fact, this behavior can be seen from the expression of asymptotic variance derived in (3.8). However, it will be evident from our simulation of Section 4 below that θ_N may possesses smaller bias and mean squared error (MSE) than those of $\widetilde{\theta}$ for finite sample sizes and, at the same time, θ_N would be much more robust than $\widetilde{\theta}$.

We now compare the performance of the MHD estimator θ_N defined at (2.6) with Zhang's [11] maximum semiparametric likelihood estimator $\widetilde{\theta} = (\widetilde{\alpha}, \widetilde{\beta})$ by examining their biases, MSEs and α-IFs. In our simulation, we let $\mu = 0.5$ be fixed and therefore $\theta = (\alpha, \beta) = (-0.125, 0.5)$. For each pair (n, m), we generated 500 independent sets of combined random samples of size $N = n + m = 60$ from the $N(0, 1)$ and $N(\mu, 1)$ distributions. Here the pair (n, m) takes varying values $(10, 50)$, $(20, 40)$, $(30, 30)$, $(40, 20)$ and $(50, 10)$. For each pair considered, we obtained estimates of the bias and MSE as follows:

$$\widehat{\text{Bias}} = \frac{1}{N_s}\sum_{i=1}^{N_s}(\widehat{\gamma}_i - \gamma)$$

and

$$\widehat{\text{MSE}} = \frac{1}{N_s}\sum_{i=1}^{N_s}(\widehat{\gamma}_i - \gamma)^2,$$

where N_s is the number of replications ($N_s = 500$ in our case), and $\widehat{\gamma}_i$ denotes an estimate of γ for the ith replication. Here $\gamma = \alpha$ or β, and $\widehat{\gamma}$ denotes either the proposed MHD estimators $\widehat{\alpha}$ and $\widehat{\beta}$ in (2.6), or the maximum semiparametric likelihood estimators $\widetilde{\alpha}$ and $\widetilde{\beta}$ of [11]. The bandwidths b_n and b_m in (2.2) and (2.3), respectively, were taken to be $h_n = n^{-2/5}$ and $h_m = m^{-2/5}$. We used Epanechnikov kernel function given by

$$K(x) = \frac{3}{4}\left(1 - x^2\right)I_{[-1,1]}(x), \tag{4.1}$$

for both K_0 and K_1. As discussed in Remark 3.3, the above choices of kernels and bandwidths satisfy conditions (C0)–(C10) and (C5′), and therefore Theorem 3.2 holds. Our simulation results are summarized in Table 2. From Table 2, it is clear that for (n, m) values $(40, 20)$ and $(50, 10)$, $\widehat{\alpha}$ is better than $\widetilde{\alpha}$ when estimated biases and MSEs are compared. For $(20, 40)$, $\widehat{\beta}$ has a smaller estimated bias than that of $\widetilde{\beta}$.

In our simulation, we also examined the behavior of the MHD estimator when data-driven bandwidths are employed. We considered the adaptive kernel density estimators mentioned in Section 2, i.e., b_n and b_m are replaced with $S_n b_n$ and $S_m b_m$, respectively, where S_n and S_m are some robust scale statistics. We used the following robust scale estimators proposed by [30],

$$S_n = 1.1926\ \text{med}_i\big(\text{med}_j(|X_i - X_j|)\big)$$
$$S_m = 1.1926\ \text{med}_i\big(\text{med}_j(|Z_i - Z_j|)\big).$$

Table 2
Estimates of the biases and MSEs of $\theta_N = (\widehat{\alpha}, \widehat{\beta})$ and $\widetilde{\theta} = (\widetilde{\alpha}, \widetilde{\beta})$ defined in (2.6) and [11], respectively, when g and h are the densities of $N(0, 1)$ and $N(0.5, 1)$, respectively.

(n, m)	$\widehat{\text{Bias}}(\widehat{\alpha})$	$\widehat{\text{MSE}}(\widehat{\alpha})$	$\widehat{\text{Bias}}(\widehat{\beta})$	$\widehat{\text{MSE}}(\widehat{\beta})$	$\widehat{\text{Bias}}(\widetilde{\alpha})$	$\widehat{\text{MSE}}(\widetilde{\alpha})$	$\widehat{\text{Bias}}(\widetilde{\beta})$	$\widehat{\text{MSE}}(\widetilde{\beta})$
(10, 50)	0.086 (0.089)	0.023 (0.024)	−0.036 (−0.050)	0.117 (0.112)	0.060	0.022	0.034	0.105
(20, 40)	0.046 (0.048)	0.012 (0.013)	−0.036 (−0.044)	0.091 (0.090)	0.013	0.011	0.046	0.085
(30, 30)	0.033 (0.035)	0.009 (0.009)	−0.050 (−0.056)	0.075 (0.074)	0.006	0.010	0.015	0.066
(40, 20)	0.012 (0.013)	0.016 (0.016)	−0.042 (−0.044)	0.094 (0.094)	−0.019	0.018	0.033	0.087
(50, 10)	−0.009 (−0.007)	0.020 (0.020)	−0.065 (−0.068)	0.132 (0.129)	−0.031	0.024	−0.010	0.126

For the same samples used to produce Table 2, the corresponding estimates of $\widehat{\alpha}$ and $\widehat{\beta}$ are also presented in Table 2: the values in the brackets in each cell are MHD estimates when adaptive kernels are employed. It appears that the adaptive kernel density estimators have not much improved the performance of the MHD estimators, and the conclusion reached from the Monte Carlo study has not impacted by bandwidth choice.

In order to examine the robustness of the estimators θ_N and $\widetilde{\theta}$, we examined their behaviors in the presence of a single outlying observation. For this purpose, the α-IF given in [12] is a suitable measure of change in estimators. Here we have used the adapted version of the α-IF applied by [26], among many others. Note that the outlier may arise from either $g(x)$ or $h(x)$. We only considered the case that the outlying value is from $h(x)$, and similar results apply to the other case as well. After drawing two data sets of the specified sizes n and m, we replaced the last observation obtained from density $h(x)$ by an integer between −9 and 11. The contamination rate is then 1/60 and the α-IFs are calculated by averaging the function

$$IF(x) = \frac{W\big((X_i)_{i=1}^n, (x, Z_i)_{i=1}^{m-1}\big) - W\big((X_i)_{i=1}^n, (Z_i)_{i=1}^m\big)}{1/60},$$

over 500 replications, where W represents any functional (estimator of θ) based on the data sets from $g(x)$ and $h(x)$. In the present situation, W is either θ_N or $\widetilde{\theta}$. For 500 replications, the α-IFs for different pairs of (n, m) are displayed in Fig. 1. The preceding figure is a clear evidence of better robustness properties of θ_N than $\widetilde{\theta}$ in the sense of resistance to a single outlying observation.

It can be seen from Fig. 1 that as the outlier increases in its absolute value, the α-IFs of θ_N (solid and dashed lines) appear to converge to constants. In fact, the absolute values of the α-IFs of θ_N reach their peaks when outlying observation is around −1 and then slide down to 0 baseline on both directions with a constant outside the interval [−5, 5]. For $\widetilde{\theta}$, however, its α-IF increases dramatically in absolute value when the outlying observation moves to left from −1. When the outlier is bigger than −1, θ_N and $\widetilde{\theta}$ are competitive but $\widetilde{\theta}$ still has larger α-IF in absolute value than θ_N. The behavior of the α-IF of $\widetilde{\theta}$ could be expected from the fact that the semiparametric likelihood is proportional in some sense to the quantity $\prod_{i=1}^m \frac{\exp[\alpha+\beta Z_i]}{n+m\exp[\alpha+\beta Z_i]}$. Without an outlying observation, $\widetilde{\beta}$ should be a value around $\beta = 0.5$. When the outlying observation x is a positive large value, $\frac{\exp[\widetilde{\alpha}+\widetilde{\beta}x]}{n+m\exp[\widetilde{\alpha}+\widetilde{\beta}x]}$ is not an extremely small value and therefore $\widetilde{\beta}$ is not much affected. If x is a negative value with $|x|$ large enough, then $\frac{\exp[\widetilde{\alpha}+\widetilde{\beta}x]}{n+m\exp[\widetilde{\alpha}+\widetilde{\beta}x]}$ will be extremely small and hence the maximizing process will tend to assign $\widetilde{\beta}$ a negative value with a large absolute value. Therefore, when x is negative with $|x|$ large enough, then the α-IF will be negative with large absolute values as shown in Fig. 1.

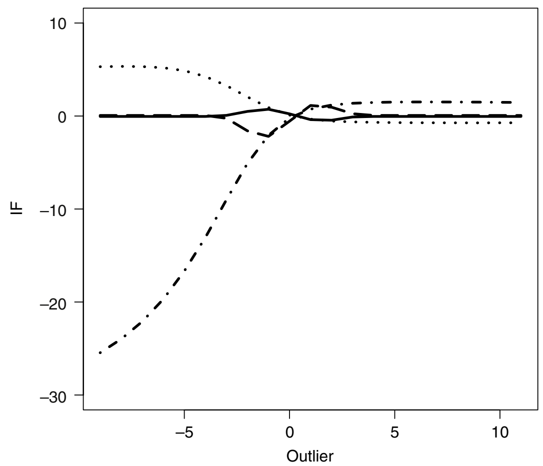
5. An example

On the basis of data from 100 participants, [31] studies the relationship between age and coronary disease status. Table 3 lists the values for age (X) and presence of evidence of significant heart disease ($Y = 1$: "*Yes*", $Y = 0$: "*No*"). Then the sample data (X_i, Y_i), $i = 1, \ldots, 100$, can be thought of as being drawn independently and identically from the joint distribution of (X, Y). The proposed MHD estimate can be applied to this data set with $n = 57$ and $m = 43$. The bandwidths were chosen as $h_n = n^{-2/5}$ and $h_m = m^{-2/5}$ and the Epanechnikov kernel function defined in (4.1) is employed for the two kernels K_0 and K_1 in (2.2) and (2.3), respectively. By fitting the model (1.1), we obtained the estimates for θ as $\theta_N = (\widehat{\alpha}, \widehat{\beta}) = (-4.64, 0.09)$. When compared with the estimates given in [11], $(\widetilde{\alpha}, \widetilde{\beta}) = (-5.03, 0.11)$, our estimates seem more conservative, i.e., they are smaller in absolute values than those in [11].

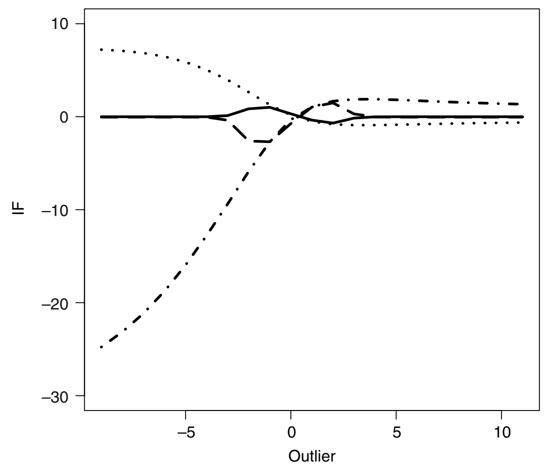
To compare the robustness of the MHD estimator, θ_N, and the maximum semiparametric likelihood estimator, $\widetilde{\theta}$, we contaminated the data and observed the change in behavior of θ_N and $\widetilde{\theta}$. Two observations were replaced: (20, 0) by (10, 1) and (69, 1) by (99, 0). The resulting MHD estimates remained unchanged, whereas the maximum semiparametric likelihood estimates were significantly affected and ended up with $\widetilde{\theta} = (\widetilde{\alpha}, \widetilde{\beta}) = (-3.16, 0.07)$. Thus, in this example, the MHD estimator is clearly more robust. This is another evidence of the fact that the MHD estimator is more resistant against outliers than the maximum semiparametric likelihood estimator.

6. Proof of asymptotic normality

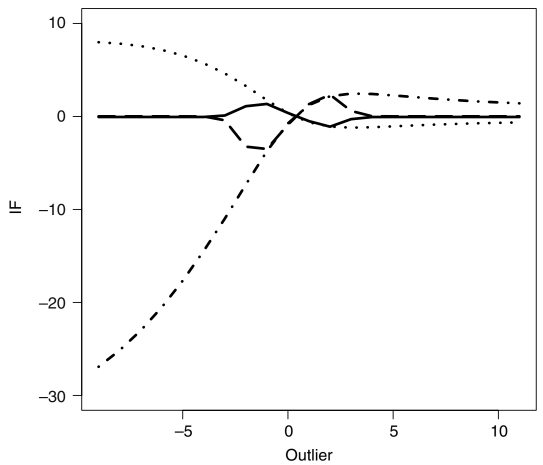
To prove Theorem 3.2, we first state a series of lemmas that are employed in the proof.



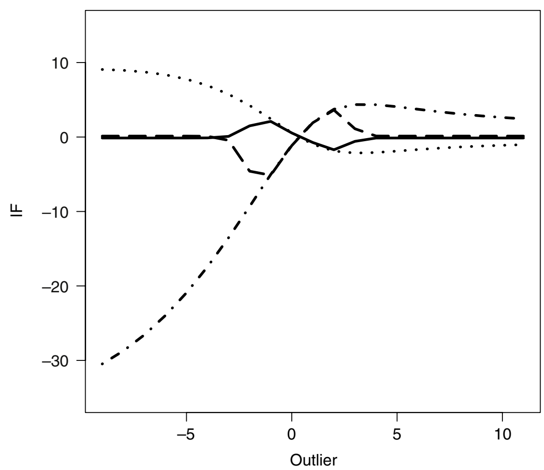
(a) $(n, m) = (10, 50)$.



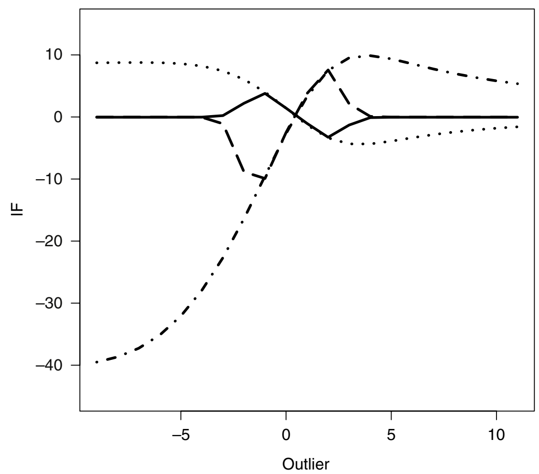
(b) $(n, m) = (20, 40)$.



(c) $(n, m) = (30, 30)$.



(d) $(n, m) = (40, 20)$.



(e) $(n, m) = (50, 10)$.

Fig. 1. The α-influence functions for $\widehat{\alpha}$ (solid), $\widehat{\beta}$ (dashed), $\widetilde{\alpha}$ (dotted) and $\widetilde{\beta}$ (dot-dashed) with respect to single outlier, where $\theta_N = (\widehat{\alpha}, \widehat{\beta})$ and $\widetilde{\theta} = (\widetilde{\alpha}, \widetilde{\beta})$ are defined in (2.6) and [11], respectively.

Table 3
Age and coronary heart disease status (CHD) of 100 subjects.

AGE	CHD	AGE	CHD	AGE	CHD	AGE	CHD	AGE	CHD	AGE	CHD	AGE	CHD
20	0	30	1	37	0	43	0	48	1	55	1	60	0
23	0	32	0	37	1	43	0	48	1	56	1	60	1
24	0	32	0	37	0	43	1	49	0	56	1	61	1
25	0	33	0	38	0	44	0	49	0	56	1	62	1
25	1	33	0	38	0	44	0	49	1	57	0	62	1
26	0	34	0	39	0	44	1	50	0	57	0	63	1
26	0	34	0	39	1	44	1	50	1	57	1	64	0
28	0	34	1	40	0	45	0	51	0	57	1	64	1
28	0	34	0	40	1	45	1	52	0	57	1	65	1
29	0	34	0	41	0	46	0	52	1	57	1	69	1
30	0	35	0	41	0	46	1	53	1	58	0		
30	0	35	0	42	0	47	0	53	1	58	1		
30	0	36	0	42	0	47	0	54	1	58	1		
30	0	36	1	42	0	47	1	55	0	59	1		
30	0	36	0	42	1	48	0	55	1	59	1		

Lemma 6.1. *Suppose that* (C3)–(C6) *hold. Then as* $N \to \infty$,

$$N^{1/2} \int \varepsilon_N(x) \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x) h_m^{1/2}(x) \mathrm{d}x \xrightarrow{P} 0, \tag{6.1}$$

$$N^{1/2} \int \varepsilon_N(x) \exp\left[\frac{1}{2}(1, r(x))\theta\right] h_\theta^{1/2}(x) g_n^{1/2}(x) \mathrm{d}x \xrightarrow{P} 0. \tag{6.2}$$

Proof. By Cauchy–Schwarz Inequality,

$$\begin{aligned} & N \cdot E\left[\int \varepsilon_{Ni}(x) \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x) h_m^{1/2}(x) \mathrm{d}x\right]^2 \\ & \quad \le N \cdot E\left[\int \varepsilon_{Ni}^2(x) \exp[(1, r(x))\theta] g_n(x) \mathrm{d}x\right] \cdot E\left[\int I_{\{|x| > \alpha_N\}} h_m(x) \mathrm{d}x\right] \\ & \quad = N \cdot \Delta_1 \cdot \Delta_2, \quad \text{say}. \end{aligned}$$

Note that by a Taylor expansion and using assumptions (C4) and (C5)

$$\begin{aligned} |\Delta_1| &= \int\int \varepsilon_{Ni}^2(x) \exp[(1, r(x))\theta] \frac{1}{b_n} K_0\left(\frac{y-x}{b_n}\right) g(y) \mathrm{d}y \mathrm{d}x \\ &= \int \varepsilon_{Ni}^2(x) \exp[(1, r(x))\theta] \int_{-a_0}^{a_0} K_0(t) g(x + t b_n) \mathrm{d}t \mathrm{d}x \\ &= \int \varepsilon_{Ni}^2(x) \exp[(1, r(x))\theta] \int_{-a_0}^{a_0} K_0(t) \left(g(x) + g^{(1)}(x) t b_n + \frac{1}{2} g^{(2)}(\xi) t^2 b_n^2\right) \mathrm{d}t \mathrm{d}x \\ &\le \int r_i^2(x) h_\theta(x) \mathrm{d}x + \frac{1}{2} b_n^2 \int \varepsilon_{Ni}^2(x) h_\theta(x) \sup_{|t| \le a_0} \frac{|g^{(2)}(x + t b_n)|}{g(x)} \mathrm{d}x \int_{-a_0}^{a_0} t^2 K_0(t) \mathrm{d}t \\ &= O(1), \end{aligned}$$

i.e., Δ_1 is bounded. On the other hand,

$$\begin{aligned} |\Delta_2| &= \int\int I_{\{|x| > \alpha_N\}} \frac{1}{b_m} K_1\left(\frac{y-x}{b_m}\right) h_\theta(y) \mathrm{d}y \mathrm{d}x \\ &= \int\int I_{\{|x| > \alpha_N\}} K_1(t) h_\theta(x + t b_m) \mathrm{d}t \mathrm{d}x \\ &= \int_{-a_1}^{a_1} K_1(t) \int_{|z - t b_m| > \alpha_N} h_\theta(z) \mathrm{d}z \mathrm{d}t \\ &\le \int_{-a_1}^{a_1} K_1(t) \mathrm{d}t \int_{|z| > \alpha_N - a_1 b_m} h_\theta(z) \mathrm{d}z \\ &= P(|Z_1| > \alpha_N - a_1 b_m). \end{aligned} \tag{6.3}$$

By assumption (C6) we have that $N \cdot E\left[\int \varepsilon_{Ni}(x) \exp[\frac{1}{2}(1, r(x))\theta] g_n^{1/2}(x) h_m^{1/2}(x) \mathrm{d}x\right]^2 \to 0$, i.e., (6.1) holds.

By Cauchy–Schwarz Inequality and using a similar argument as in (6.3),

$$
\begin{aligned}
& N \cdot E\left[\int \varepsilon_{Ni}(x) \exp\left[\frac{1}{2}(1, r(x))\theta\right] h_\theta^{1/2}(x) g_n^{1/2}(x)\mathrm{d}x\right]^2 \\
&\quad \le N \cdot \int r_i^2(x) \exp[(1, r(x))\theta] h_\theta(x)\mathrm{d}x \cdot E\left[\int I_{\{|x|>\alpha_N\}} g_n(x)\mathrm{d}x\right] \\
&\quad = N \cdot \int r_i^2(x) \exp[(1, r(x))\theta] h_\theta(x)\mathrm{d}x \cdot \int\int I_{\{|x|>\alpha_N\}} \frac{1}{b_n} K_0\left(\frac{y-x}{b_n}\right) g(y)\mathrm{d}y\mathrm{d}x \\
&\quad \le N \cdot \int r_i^2(x) \exp[(1, r(x))\theta] h_\theta(x)\mathrm{d}x \cdot P(|X_1| > \alpha_N - a_0 b_n),
\end{aligned}
$$

and by assumptions (C4) and (C6) we have that (6.2) holds. □

Lemma 6.2. *Suppose that* (C0)–(C3) *and* (C7) *hold. Then as* $N \to \infty$,

$$
N^{1/2} \int |\delta_N(x)| \left(h_m^{1/2}(x) - h_\theta^{1/2}(x)\right)^2 \mathrm{d}x \xrightarrow{P} 0, \tag{6.4}
$$

$$
N^{1/2} \int |\delta_N(x)| \exp[(1, r(x))\theta] \left(g_n^{1/2}(x) - g^{1/2}(x)\right)^2 \mathrm{d}x \xrightarrow{P} 0. \tag{6.5}
$$

Proof. Note that

$$
\begin{aligned}
N^{1/2} \int |\delta_N(x)| \left(h_m^{1/2}(x) - h_\theta^{1/2}(x)\right)^2 \mathrm{d}x &\le N^{1/2} \int |\delta_N(x)| h_\theta^{-1}(x) \left(h_m(x) - h_\theta(x)\right)^2 \mathrm{d}x \\
&\le 2\Big[N^{1/2} \int |\delta_N(x)| h_\theta^{-1}(x) \left(h_m(x) - Eh_m(x)\right)^2 \mathrm{d}x \\
&\quad + N^{1/2} \int |\delta_N(x)| h_\theta^{-1}(x) \left(Eh_m(x) - h_\theta(x)\right)^2 \mathrm{d}x\Big] \\
&= 2(A_{1N} + A_{2N}), \quad \text{say.}
\end{aligned}
$$

By conditions (C0), (C2), (C3) and (C7) as $N \to \infty$,

$$
\begin{aligned}
E|A_{1N}| &= N^{1/2} \int |\delta_N(x)| h_\theta^{-1}(x) E\left(h_m(x) - Eh_m(x)\right)^2 \mathrm{d}x \\
&\le N^{1/2} \int |\delta_N(x)| h_\theta^{-1}(x) \frac{1}{m b_m^2} \int K_1^2\left(\frac{y-x}{b_m}\right) h_\theta(y)\mathrm{d}y\mathrm{d}x \\
&= N^{1/2} m^{-1} b_m^{-1} \int |\delta_N(x)| \int_{-a_1}^{a_1} K_1^2(t) h_\theta(x + t b_m) h_\theta^{-1}(x)\mathrm{d}t\mathrm{d}x \\
&\le N^{1/2} m^{-1} b_m^{-1} \int |\delta_N(x)| \sup_{|t|\le a_1} \frac{h_\theta(x + t b_m)}{h_\theta(x)} \mathrm{d}x \int_{-a_1}^{a_1} K_1^2(t)\mathrm{d}t \\
&\to 0,
\end{aligned}
$$

i.e., $A_{1N} \xrightarrow{P} 0$ as $N \to \infty$. By a Taylor expansion and using conditions (C1) and (C7),

$$
\begin{aligned}
|A_{2N}| &= N^{1/2} \int |\delta_N(x)| h_\theta^{-1}(x) \Big[\int_{-a_1}^{a_1} K_1(t)\left(h_\theta(x + t b_m) - h_\theta(x)\right)\mathrm{d}t\Big]^2 \mathrm{d}x \\
&\le \frac{1}{4} N^{1/2} b_m^4 \int |\delta_N(x)| h_\theta^{-1}(x) \Big[\sup_{|t|\le a_1} |h_\theta^{(2)}(x + t b_m)| \int_{-a_1}^{a_1} t^2 K_1(t)\mathrm{d}t\Big]^2 \mathrm{d}x \\
&\le \frac{1}{4} N^{1/2} b_m^4 \int |\delta_N(x)| h_\theta(x) \sup_{|t|\le a_1} \left[\frac{h_\theta^{(2)}(x + t b_m)}{h_\theta(x)}\right]^2 \mathrm{d}x \left(\int_{-a_1}^{a_1} t^2 K_1(t)\mathrm{d}t\right)^2 \\
&\to 0.
\end{aligned}
$$

Hence (6.4) holds. Proof of (6.5) is similar to that of (6.4). □

Lemma 6.3. *Suppose that* (C0)–(C7) *hold. Then the asymptotic distribution of*

$$
N^{1/2} \int (1, r(x))^T \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x) \left(h_m^{1/2}(x) - h_\theta^{1/2}(x)\right) \mathrm{d}x \tag{6.6}
$$

is the same as that of

$$N^{1/2}\int \delta_N(x)h_\theta^{1/2}(x)\left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)\mathrm{d}x.$$

Proof. From Lemma 6.1,

$$N^{1/2}\int \varepsilon_N(x)\exp\left[\frac{1}{2}(1,r(x))\theta\right]g_n^{1/2}(x)\left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)\mathrm{d}x\xrightarrow{P}0,$$

and as a result the asymptotic distribution of (6.6) is the same as that of

$$N^{1/2}\int \delta_N(x)\exp\left[\frac{1}{2}(1,r(x))\theta\right]g_n^{1/2}(x)\left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)\mathrm{d}x.$$

By Cauchy–Schwarz Inequality

$$\left\{N^{1/2}\int \delta_{Ni}(x)\exp\left[\frac{1}{2}(1,r(x))\theta\right]\left(g_n^{1/2}(x)-g^{1/2}(x)\right)\left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)\mathrm{d}x\right\}^2$$
$$\leq N^{1/2}\int |\delta_{Ni}(x)|\exp[(1,r(x))\theta]\left(g_n^{1/2}(x)-g^{1/2}(x)\right)^2\mathrm{d}x\, N^{1/2}\int |\delta_{Ni}(x)|\left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)^2\mathrm{d}x,$$

which is $o_P(1)$ by Lemma 6.2. Hence the result follows. □

Remark 6.1. In fact, the asymptotic distribution of (6.6) is the same as that of

$$N^{1/2}\int (1,r(x))^T h_\theta^{1/2}(x)\left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)\mathrm{d}x.$$

The reason being that as $N\to\infty$,

$$N^{1/2}\int \varepsilon_N(x)h_\theta^{1/2}(x)\left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)\mathrm{d}x\xrightarrow{P}0$$

under conditions (C3), (C4) and (C6). The proof is similar to that of Lemma 6.1 and is therefore omitted.

Remark 6.2. Instead of condition (C7), if h_θ and g have bounded second derivatives and conditions (C9) and (C11) hold, then Lemma 6.3 still holds. Since

$$\left\{N^{1/2}\int \delta_{Ni}(x)\exp\left[\frac{1}{2}(1,r(x))\theta\right]\left(g_n^{1/2}(x)-g^{1/2}(x)\right)\left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)\mathrm{d}x\right\}^2$$
$$\leq N^{1/2}\int |\delta_{Ni}(x)|\exp\left[\frac{1}{2}(1,r(x))\theta\right]\left(g_n^{1/2}(x)-g^{1/2}(x)\right)^2\mathrm{d}x$$
$$\times N^{1/2}\int |\delta_{Ni}(x)|\exp\left[\frac{1}{2}(1,r(x))\theta\right]\left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)^2\mathrm{d}x,$$

similar arguments as in the proof of Lemmas 6.2 and 6.3 give above conclusion.

Lemma 6.4. *Suppose that* (C4) *and* (C6) *hold. Then as* $N\to\infty$,

$$N^{1/2}\int |\varepsilon_N(x)|h_\theta(x)\mathrm{d}x\to 0,$$
$$N^{1/2}\cdot\frac{1}{m}\sum_{i=1}^{m}\varepsilon_N(Z_i)\xrightarrow{P}0,$$
$$N^{1/2}\cdot\frac{1}{n}\sum_{i=1}^{n}\varepsilon_N(X_i)\exp[(1,r(X_i))\theta]\xrightarrow{P}0.$$

Proof. By Cauchy–Schwarz Inequality,

$$\begin{aligned}N^{1/2}\int |\varepsilon_{Ni}(x)|h_\theta(x)\mathrm{d}x &\leq \left[N\int I_{\{|x|>\alpha_N\}}h_\theta(x)\mathrm{d}x\right]^{1/2}\left[\int r_i^2(x)h_\theta(x)\mathrm{d}x\right]^{1/2}\\ &= \left[NP(|Z_1|>\alpha_N)\right]^{1/2}\left[\int r_i^2(x)h_\theta(x)\mathrm{d}x\right]^{1/2}\\ &\to 0.\end{aligned}$$

As a result,

$$\begin{aligned}
E\left|N^{1/2}\cdot\frac{1}{m}\sum_{i=1}^{m}\varepsilon_N(Z_i)\right| &\leq E\left[N^{1/2}\cdot\frac{1}{m}\sum_{i=1}^{m}|\varepsilon_N(Z_i)|\right]\\
&= N^{1/2}\int|\varepsilon_N(x)|h_\theta(x)\mathrm{d}x\\
&\to 0,\\
E\left|N^{1/2}\cdot\frac{1}{n}\sum_{i=1}^{n}\varepsilon_N(X_i)\exp[(1,r(X_i))\theta]\right| &\leq E\left[N^{1/2}\cdot\frac{1}{n}\sum_{i=1}^{n}|\varepsilon_N(X_i)|\exp[(1,r(X_i))\theta]\right]\\
&= N^{1/2}\int|\varepsilon_N(x)|h_\theta(x)\mathrm{d}x\\
&\to 0,
\end{aligned}$$

and hence the results follow. □

Lemma 6.5. *Suppose that* (C0)–(C4) *and* (C8)–(C10) *hold. Then as* $N\to\infty$,

$$N^{1/2}\int\delta_N(x)h_m(x)\mathrm{d}x-N^{1/2}\frac{1}{m}\sum_{i=1}^{m}\delta_N(Z_i)\xrightarrow{P}0,$$

$$N^{1/2}\int\delta_N(x)\exp[(1,r(x))\theta]g_n(x)\mathrm{d}x-N^{1/2}\frac{1}{n}\sum_{i=1}^{n}\delta_N(X_i)\exp[(1,r(X_i))\theta]\xrightarrow{P}0.$$

Proof. We give only the proof for the second convergence, and the proof for the first convergence is similar. Let $D_{Ni}=N^{1/2}\int\delta_{Ni}(x)\exp[(1,r(x))\theta]g_n(x)\mathrm{d}x-N^{1/2}\frac{1}{n}\sum_{i=1}^{n}\delta_{Ni}(X_i)\exp[(1,r(X_i))\theta]$, $i=0,1,\ldots,p$. Then by (C8)

$$\begin{aligned}
|E[D_{Ni}]| &= N^{1/2}\left|\int\delta_{Ni}(x)\exp[(1,r(x))\theta]E[g_n(x)]\mathrm{d}x-\int\delta_{Ni}(x)h_\theta(x)\mathrm{d}x\right|\\
&= N^{1/2}\left|\int\delta_{Ni}(x)\exp[(1,r(x))\theta]\int_{-a_0}^{a_0}K_0(t)\big(g(x+tb_n)-g(x)\big)\mathrm{d}t\mathrm{d}x\right|\\
&\leq N^{1/2}b_n^2\int|\delta_{Ni}(x)|h_\theta(x)\sup_{|t|\leq a_0}\frac{|g^{(2)}(x+tb_n)|}{g(x)}\mathrm{d}x\int_{-a_0}^{a_0}t^2K_0(t)\mathrm{d}t\\
&\to 0.
\end{aligned}$$

Note that

$$\begin{aligned}
Var[D_{Ni}] &\leq \frac{N}{n}E\left[\int\delta_{Ni}(x)\exp[(1,r(x))\theta]\frac{1}{b_n}K_0\left(\frac{x-X_1}{b_n}\right)\mathrm{d}x-\delta_{Ni}(X_1)\exp[(1,r(X_1))\theta]\right]^2\\
&= \frac{N}{n}E\left[\int_{-a_0}^{a_0}K_0(t)\Big(\delta_{Ni}(X_1+tb_n)\exp[(1,r(X_1+tb_n))\theta]-\delta_{Ni}(X_1)\exp[(1,r(X_1))\theta]\Big)\mathrm{d}t\right]^2\\
&= \frac{N}{n}E\left[\int_{-a_0}^{a_0}K_0(t)r_i(X_1+tb_n)\exp[(1,r(X_1+tb_n))\theta]\Big(I_{\{|X_1+tb_n|\leq\alpha_N\}}-I_{\{|X_1|\leq\alpha_N\}}\Big)\mathrm{d}t\right.\\
&\quad\left.+\int_{-a_0}^{a_0}K_0(t)I_{\{|X_1|\leq\alpha_N\}}\Big(r_i(X_1+tb_n)\exp[(1,r(X_1+tb_n))\theta]-r_i(X_1)\exp[(1,r(X_1))\theta]\Big)\mathrm{d}t\right]^2\\
&\leq \frac{2N}{n}\left\{E\left[\int_{-a_0}^{a_0}K_0(t)r_i(X_1+tb_n)\exp[(1,r(X_1+tb_n))\theta]\Big(I_{\{|X_1+tb_n|\leq\alpha_N\}}-I_{\{|X_1|\leq\alpha_N\}}\Big)\mathrm{d}t\right]^2\right.\\
&\quad\left.+E\left[\int_{-a_0}^{a_0}K_0(t)I_{\{|X_1|\leq\alpha_N\}}\Big(r_i(X_1+tb_n)\exp[(1,r(X_1+tb_n))\theta]-r_i(X_1)\exp[(1,r(X_1))\theta]\Big)\mathrm{d}t\right]^2\right\}\\
&= \frac{2N}{n}(B_{Ni}+C_{Ni}),\quad\text{say.}
\end{aligned}$$

By Cauchy–Schwarz Inequality,

$$B_{Ni}\leq E\int_{-a_0}^{a_0}K_0(t)r_i^2(X_1+tb_n)\exp[2(1,r(X_1+tb_n))\theta]\big(I_{\{|X_1+tb_n|\leq\alpha_N\}}-I_{\{|X_1|\leq\alpha_N\}}\big)^2\mathrm{d}t$$

$$
\begin{aligned}
&= \int_0^{a_0} K_0(t)\Bigg[\int_{-\alpha_N-tb_n}^{-\alpha_N} r_i^2(y+tb_n)\exp[2(1,r(y+tb_n))\theta]g(y)\mathrm{d}y \\
&\quad + \int_{\alpha_N-tb_n}^{\alpha_N} r_i^2(y+tb_n)\exp[2(1,r(y+tb_n))\theta]g(y)\mathrm{d}y\Bigg]\mathrm{d}t \\
&\quad + \int_{-a_0}^{0} K_0(t)\Bigg[\int_{-\alpha_N}^{-\alpha_N-tb_n} r_i^2(y+tb_n)\exp[2(1,r(y+tb_n))\theta]g(y)\mathrm{d}y \\
&\quad + \int_{\alpha_N}^{\alpha_N-tb_n} r_i^2(y+tb_n)\exp[2(1,r(y+tb_n))\theta]g(y)\mathrm{d}y\Bigg]\mathrm{d}t.
\end{aligned} \tag{6.7}
$$

Note that $r_i^2(x)\exp[(1,r(x))\theta]h_\theta(x)$ is bounded by (C4) and therefore by (C9)

$$
\begin{aligned}
&\int_0^{a_0} K_0(t)\int_{-\alpha_N-tb_n}^{-\alpha_N} r_i^2(y+tb_n)\exp[2(1,r(y+tb_n))\theta]g(y)\mathrm{d}y \\
&= \int_0^{a_0} K_0(t)\int_{-\alpha_N}^{-\alpha_N+tb_n} r_i^2(y)\exp[2(1,r(y))\theta]g(y-tb_n)\mathrm{d}y\mathrm{d}t \\
&\le \sup_{|x|\le\alpha_N}\sup_{|t|\le a_0}\frac{g(x+tb_n)}{g(x)}\int_0^{a_0} K_0(t)\int_{-\alpha_N}^{-\alpha_N+tb_n} r_i^2(y)\exp[(1,r(y))\theta]h_\theta(y)\mathrm{d}y\mathrm{d}t \\
&= O\left(b_n\int_0^{a_0} tK_0(t)\mathrm{d}t\right) \\
&\to 0,
\end{aligned}
$$

as $N\to\infty$, and other three terms on the r.h.s. of (6.7) go to zero using similar arguments. Thus $B_{Ni}\to 0$ as $N\to\infty$. For C_{Ni}, by Cauchy–Schwarz inequality and (C10) we have

$$
\begin{aligned}
C_{Ni} &\le E\left[\int_{-a_0}^{a_0} K_0(t)I_{\{|X_1|\le\alpha_N\}}\Big(r_i(X_1+tb_n)\exp[(1,r(X_1+tb_n))\theta]-r_i(X_1)\exp[(1,r(X_1))\theta]\Big)^2\mathrm{d}t\right] \\
&= \int_{-a_0}^{a_0} K_0(t)\int I_{\{|x|\le\alpha_N\}}\Big(r_i(x+tb_n)\exp[(1,r(x+tb_n))\theta]-r_i(x)\exp[(1,r(x))\theta]\Big)^2 g(x)\mathrm{d}x\mathrm{d}t \\
&\le b_n^2\int I_{\{|x|\le\alpha_N\}}g(x)\sup_{|t|\le a_0}\left[\frac{\partial r_i(y)\exp[(1,r(y))\theta]}{\partial y}\Big|_{y=x+tb_n}\right]^2\mathrm{d}x\int_{-a_0}^{a_0} t^2K_0(t)\mathrm{d}t \\
&\to 0.
\end{aligned}
$$

Thus $Var[D_{Ni}]\to 0$ as $N\to\infty$. This yields that $E[D_{Ni}^2]=Var[D_{Ni}]+(E[D_{Ni}])^2\to 0$, and therefore $D_{Ni}\xrightarrow{P} 0$ as $N\to\infty$. □

Proposition 6.1. *Suppose that* (C0)–(C10) *hold. Then the asymptotic distribution of* (6.6) *is* $N\big(0,\frac{1}{4(1-\rho)}\Sigma_1\big)$ *with* Σ_1 *defined by* (3.7).

Proof. In view of Lemma 6.3, we only need to give the asymptotic distribution of $N^{1/2}\int\delta_N(x)h_\theta^{1/2}(x)\big(h_m^{1/2}(x)-h_\theta^{1/2}(x)\big)\mathrm{d}x$. Applying the following algebraic expression, with $b\ge 0$, $a>0$,

$$
b^{1/2}-a^{1/2}=\frac{b-a}{2a^{1/2}}-\frac{\big(b^{1/2}-a^{1/2}\big)^2}{2a^{1/2}}, \tag{6.8}
$$

we have that as $N\to\infty$,

$$
\begin{aligned}
&N^{1/2}\int\delta_N(x)h_\theta^{1/2}(x)\big(h_m^{1/2}(x)-h_\theta^{1/2}(x)\big)\mathrm{d}x \\
&= \frac{1}{2}N^{1/2}\int\delta_N(x)\big(h_m(x)-h_\theta(x)\big)\mathrm{d}x-\frac{1}{2}N^{1/2}\int\delta_N(x)\big(h_m^{1/2}(x)-h_\theta^{1/2}(x)\big)^2\mathrm{d}x \\
&= \frac{1}{2}N^{1/2}\int\delta_N(x)\big(h_m(x)-h_\theta(x)\big)\mathrm{d}x+o_P(1)\quad(\text{by Lemma 6.2}) \\
&= \frac{1}{2}N^{1/2}\left[\frac{1}{m}\sum_{i=1}^m\delta_N(Z_i)-\int\delta_N(x)h_\theta(x)\mathrm{d}x\right]+\frac{1}{2}N^{1/2}\left[\int\delta_N(x)h_m(x)\mathrm{d}x-\frac{1}{m}\sum_{i=1}^m\delta_N(Z_i)\right]+o_P(1)
\end{aligned}
$$

$$= \frac{1}{2}N^{1/2}\left[\frac{1}{m}\sum_{i=1}^{m}\delta_N(Z_i) - \int \delta_N(x)h_\theta(x)\mathrm{d}x\right] + o_P(1) \quad (\text{by Lemma 6.5})$$
$$= \frac{1}{2}N^{1/2}\left[\frac{1}{m}\sum_{i=1}^{m}(1, r(Z_i))^T - \int (1, r(x))^T h_\theta(x)\mathrm{d}x\right] + o_P(1) \quad (\text{by Lemma 6.4}).$$

Obviously the asymptotic distribution of $m^{1/2}\left[\frac{1}{m}\sum_{i=1}^{m}(1, r(Z_i))^T - \int(1, r(x))^T h_\theta(x)\mathrm{d}x\right]$ is $N\big(0, \Sigma_1\big)$. Hence the result. □

Lemma 6.6. *Suppose that* (C0)–(C7) *and* (C5′) *hold. Then the asymptotic distribution of*

$$N^{1/2}\int (1, r(x))^T \exp[(1, r(x))\theta]g_n^{1/2}(x)\big(g_n^{1/2}(x) - g^{1/2}(x)\big)\mathrm{d}x \tag{6.9}$$

is the same as that of

$$N^{1/2}\int \delta_N(x)\exp[(1, r(x))\theta]g^{1/2}(x)\big(g_n^{1/2}(x) - g^{1/2}(x)\big)\mathrm{d}x.$$

Proof. Note that by Cauchy–Schwarz Inequality, a Taylor expansion, (C5′) and Lemma 6.4,

$$E\left|N^{1/2}\int \varepsilon_{Ni}(x)\exp[(1, r(x))\theta]g_n(x)\mathrm{d}x\right|$$
$$\le N^{1/2}\int |\varepsilon_{Ni}(x)|\exp[(1, r(x))\theta]\int_{-a_0}^{a_0} K_0(t)g(x+tb_n)\mathrm{d}t\mathrm{d}x$$
$$\le N^{1/2}\int |\varepsilon_{Ni}(x)|\exp[(1, r(x))\theta]\int_{-a_0}^{a_0} K_0(t)\Big(g(x) + g^{(1)}(x)tb_n + \frac{1}{2}t^2b_n^2\sup_{|t|\le a_0}|g^{(2)}(x+tb_n)|\Big)\mathrm{d}t\mathrm{d}x$$
$$\le N^{1/2}\int |\varepsilon_{Ni}(x)|h_\theta(x)\mathrm{d}x + \frac{1}{2}N^{1/2}b_n^2\int |\varepsilon_{Ni}(x)|h_\theta(x)\sup_{|t|\le a_0}\frac{|g^{(2)}(x+tb_n)|}{g(x)}\int_{-a_0}^{a_0} t^2K_0(t)\mathrm{d}t$$
$$\to 0.$$

Thus $N^{1/2}\int \varepsilon_N(x)\exp[(1, r(x))\theta]g_n(x)\mathrm{d}x \xrightarrow{P} 0$. Combined with the result in Lemma 6.1, we therefore have

$$N^{1/2}\int \varepsilon_N(x)\exp[(1, r(x))\theta]g_n^{1/2}(x)\big(g_n^{1/2}(x) - g^{1/2}(x)\big)\mathrm{d}x \xrightarrow{P} 0,$$

and so the asymptotic distribution of (6.9) is the same as that of

$$N^{1/2}\int \delta_N(x)\exp[(1, r(x))\theta]g_n^{1/2}(x)\big(g_n^{1/2}(x) - g^{1/2}(x)\big)\mathrm{d}x \xrightarrow{P} 0.$$

The result now follows from Lemma 6.2. □

Proposition 6.2. *Suppose that* (C0)–(C10) *and* (C5′) *hold. Then the asymptotic distribution of* (6.9) *is* $N\big(0, \frac{1}{4\rho}\Sigma_0\big)$ *with* Σ_0 *defined by* (3.6).

Proof. Similar to that of Proposition 6.1.

Again in view of Lemma 6.6, we only need to give the asymptotic distribution of $N^{1/2}\int \delta_N(x)\exp[(1, r(x))\theta]g^{1/2}(x)\big(g_n^{1/2}(x) - g^{1/2}(x)\big)\mathrm{d}x$. Applying the algebraic expression (6.8) we have that as $N\to\infty$,

$$N^{1/2}\int \delta_N(x)\exp[(1, r(x))\theta]g^{1/2}(x)\big(g_n^{1/2}(x) - g^{1/2}(x)\big)\mathrm{d}x$$
$$= \frac{1}{2}N^{1/2}\int \delta_N(x)\exp[(1, r(x))\theta]\big(g_n(x) - g(x)\big)\mathrm{d}x + \frac{1}{2}N^{1/2}\int \delta_N(x)\exp[(1, r(x))\theta]\big(g_n^{1/2}(x) - g^{1/2}(x)\big)^2\mathrm{d}x$$
$$= \frac{1}{2}N^{1/2}\int \delta_N(x)\exp[(1, r(x))\theta]\big(g_n(x) - g(x)\big)\mathrm{d}x + o_P(1) \quad (\text{by Lemma 6.2})$$
$$= \frac{1}{2}N^{1/2}\Big\{\frac{1}{n}\sum_{i=1}^{n}\delta_N(X_i)\exp[(1, r(X_i))\theta] - \int \delta_N(x)h_\theta(x)\mathrm{d}x\Big\}$$
$$+ \frac{1}{2}N^{1/2}\Big\{\int \delta_N(x)\exp[(1, r(x))\theta]g_n(x)\mathrm{d}x - \frac{1}{n}\sum_{i=1}^{n}\delta_N(X_i)\exp[(1, r(X_i))\theta]\Big\} + o_P(1)$$

$$= \frac{1}{2} N^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \delta_N(X_i) \exp[(1, r(X_i))\theta] - \int \delta_N(x) h_\theta(x) \mathrm{d}x \right\} + o_P(1) \quad \text{(by Lemma 6.5)}$$

$$= \frac{1}{2} N^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^{n} (1, r(X_i))^T \exp[(1, r(X_i))\theta] - \int (1, r(x))^T h_\theta(x) \mathrm{d}x \right\} + o_P(1) \quad \text{(by Lemma 6.4)}.$$

Obviously the asymptotic distribution of $n^{1/2}\left[\frac{1}{n}\sum_{i=1}^{n}(1, r(X_i))^T \exp[(1, r(X_i))\theta] - \int (1, r(x))^T h_\theta(x)\mathrm{d}x\right]$ is $N(0, \Sigma_0)$. Hence the result. □

Proof of Theorem 3.2. Note that by Lemmas 6.3 and 6.6

$$N^{1/2} \int \left\{ \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x) h_m^{1/2}(x) - \exp[(1, r(x))\theta] g_n(x) \right\} (1, r(x))^T \mathrm{d}x$$

$$= N^{1/2} \int (1, r(x))^T \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x) \left(h_m^{1/2}(x) - h_\theta^{1/2}(x)\right) \mathrm{d}x$$

$$- N^{1/2} \int (1, r(x))^T \exp[(1, r(x))\theta] g_n^{1/2}(x) \left(g_n^{1/2}(x) - g^{1/2}(x)\right) \mathrm{d}x$$

$$= N^{1/2} \int \delta_N(x) h_\theta^{1/2}(x) \left(h_m^{1/2}(x) - h_\theta^{1/2}(x)\right) \mathrm{d}x - N^{1/2} \int \delta_N(x) \exp[(1, r(x))\theta] g^{1/2}(x) \left(g_n^{1/2}(x) - g^{1/2}(x)\right) \mathrm{d}x + o_P(1)$$

and the first two terms on the r.h.s. of the preceding expression are independent. Then by Propositions 6.1 and 6.2 and Slutsky's theorem, the result follows. □

Acknowledgments

We wish to thank the referees for their very constructive and helpful comments that helped to improve the paper. This paper is a part of the first author's Ph.D. thesis, which is the co-winner of the Pierre Robillard Award for the year 2007 awarded by the Statistical Society of Canada each year to the author of the best doctoral thesis in probability or statistics defended at a Canadian university (http://www.ssc.ca/main/about/awards/awards2008_e.html#Robillard). This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Appendix

Proof of Theorem 3.1. From Theorem 2.2 we have that $\theta_N \xrightarrow{P} \theta$ as $N \to \infty$. Since $t = \theta_N \in \Theta$ minimizes the Hellinger distance between $\widehat{h}_t$ and h_m, θ_N maximizes $\int \widehat{h}_t^{1/2}(x) h_m^{1/2}(x)\mathrm{d}x - \int \frac{1}{2}\widehat{h}_t(x)\mathrm{d}x$. Also since K_0 has compact support, we have $0 = \int \frac{\partial}{\partial t}[\widehat{h}_t^{1/2}(x) h_m^{1/2}(x) - \frac{1}{2}\widehat{h}_t(x)]|_{t=\theta_N}\mathrm{d}x$, i.e.,

$$\int \exp\left[\frac{1}{2}(1, r(x))\theta_N\right] g_n^{1/2}(x) h_m^{1/2}(x) (1, r(x))^T \mathrm{d}x - \int \exp[(1, r(x))\theta_N] g_n(x) (1, r(x))^T \mathrm{d}x = 0. \tag{A.1}$$

We will prove in the following that under condition (D2), (D5) or (D6), (A.1) will reduce to

$$\int \left\{ \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x) h_m^{1/2}(x) - \exp[(1, r(x))\theta] g_n(x) \right\} (1, r(x))^T \mathrm{d}x$$
$$- \left[\frac{1}{2} \int h_\theta(x) (1, r(x))^T (1, r(x)) \mathrm{d}x + c_N\right] (\theta_N - \theta) = 0, \tag{A.2}$$

where c_N is a $(p+1) \times (p+1)$ matrix with elements tending to zero in probability as $N \to \infty$, i.e., (3.4) holds.

(i) Suppose that (D2) or (D5) holds. Then for any $t \in \Theta \cap B(\theta, \epsilon)$,

$$\left| \int r_i(x) r_j(x) r_k(x) \exp[(1, r(x))t] g_n(x) \mathrm{d}x \right| \le C \int g_n(x) \mathrm{d}x = C$$

$$\left| \int r_i(x) r_j(x) r_k(x) \exp\left[\frac{1}{2}(1, r(x))t\right] g_n^{1/2}(x) h_m^{1/2}(x) \mathrm{d}x \right| \le C \left(\int g_n(x) \mathrm{d}x \right)^{1/2} \left(\int h_m(x) \mathrm{d}x \right)^{1/2} = C$$

for some positive constant C. Therefore, by a Taylor expansion of θ_N at θ, one obtains with $\theta_t = t\theta + (1-t)\theta_N$ for some $0 < t < 1$,

$$\int \exp\left[\frac{1}{2}(1, r(x))\theta_N\right] g_n^{1/2}(x) h_m^{1/2}(x) (1, r(x))^T \mathrm{d}x$$

$$= \int \left\{ \exp\left[\frac{1}{2}(1, r(x))\theta\right] + \frac{1}{2}\exp\left[\frac{1}{2}(1, r(x))\theta\right](1, r(x))(\theta_N - \theta) \right.$$
$$\left. + \frac{1}{8}\exp\left[\frac{1}{2}(1, r(x))\theta_t\right](\theta_N - \theta)^T(1, r(x))^T(1, r(x))(\theta_N - \theta) \right\} g_n^{1/2}(x)h_m^{1/2}(x)(1, r(x))^T \mathrm{d}x$$
$$= \int \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x)h_m^{1/2}(x)(1, r(x))^T \mathrm{d}x$$
$$+ \frac{1}{2}\int \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x)h_m^{1/2}(x)(1, r(x))^T(1, r(x))\mathrm{d}x(\theta_N - \theta) + a_N(\theta_N - \theta), \tag{A.3}$$

$$\int \exp[(1, r(x))\theta_N]g_n(x)(1, r(x))^T \mathrm{d}x = \int \Big\{ \exp[(1, r(x))\theta] + \exp[(1, r(x))\theta](1, r(x))(\theta_N - \theta)$$
$$+ \frac{1}{2}\exp[(1, r(x))\theta_t](\theta_N - \theta)^T(1, r(x))^T(1, r(x))(\theta_N - \theta)\Big\} g_n(x)(1, r(x))^T \mathrm{d}x$$
$$= \int \exp[(1, r(x))\theta]g_n(x)(1, r(x))^T \mathrm{d}x + \int \exp[(1, r(x))\theta]g_n(x)(1, r(x))^T(1, r(x))\mathrm{d}x(\theta_N - \theta) + b_N(\theta_N - \theta), \tag{A.4}$$

where a_N and b_N are $(p + 1) \times (p + 1)$ matrixes with elements tending to zero in probability as $N \to \infty$ by the fact that $\theta_N \to \theta$. From (A.1), (A.3) and (A.4), we obtain

$$0 = \int \left\{ \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x)h_m^{1/2}(x) - \exp[(1, r(x))\theta]g_n(x) \right\} (1, r(x))^T \mathrm{d}x$$
$$+ \left\{ \frac{1}{2}\int \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x)h_m^{1/2}(x)(1, r(x))^T(1, r(x))\mathrm{d}x \right.$$
$$\left. - \int \exp[(1, r(x))\theta]g_n(x)(1, r(x))^T(1, r(x))\mathrm{d}x \right\} (\theta_N - \theta) + [a_N - b_N](\theta_N - \theta). \tag{A.5}$$

Since either (D2) or (D5) holds,

$$\left| \int \exp\left[\frac{1}{2}(1, r(x))\theta\right] \left\{ g_n^{1/2}(x)h_m^{1/2}(x) - g^{1/2}(x)h_\theta^{1/2}(x) \right\} (1, r(x))^T(1, r(x))\mathrm{d}x \right|$$
$$\le C\left\{ \int g_n^{1/2}(x)|h_m^{1/2}(x) - h_\theta^{1/2}(x)|\mathrm{d}x + \int h_\theta^{1/2}(x)|g_n^{1/2}(x) - g^{1/2}(x)|\mathrm{d}x \right\}$$
$$\le C\left\{ \left[\int (h_m^{1/2}(x) - h_\theta^{1/2}(x))^2 \mathrm{d}x\right]^{1/2} + \left[\int (g_n^{1/2}(x) - g^{1/2}(x))^2 \mathrm{d}x\right]^{1/2} \right\}$$

with the r.h.s. of the preceding inequality goes to zero in probability using the results in Theorem 2.2. Thus,

$$\int \exp\left[\frac{1}{2}(1, r(x))\theta\right] g_n^{1/2}(x)h_m^{1/2}(x)(1, r(x))^T(1, r(x))\mathrm{d}x \xrightarrow{P} \int h_\theta(x)(1, r(x))^T(1, r(x))\mathrm{d}x. \tag{A.6}$$

Similarly

$$\left| \int \exp[(1, r(x))\theta]\big(g_n(x) - g(x)\big)(1, r(x))^T(1, r(x))\mathrm{d}x \right| \le C\int |(g_n^{1/2}(x) - g^{1/2}(x))(g_n^{1/2}(x) + g^{1/2}(x))|\mathrm{d}x$$
$$\le C\Big[\int (g_n^{1/2}(x) - g^{1/2}(x))^2 \mathrm{d}x\Big]^{1/2}\Big[\int (g_n^{1/2}(x) + g^{1/2}(x))^2 \mathrm{d}x\Big]^{1/2}$$
$$\le 2C\Big[\int (g_n^{1/2}(x) - g^{1/2}(x))^2 \mathrm{d}x\Big]^{1/2}$$
$$\xrightarrow{P} 0,$$

i.e.,

$$\int \exp[(1, r(x))\theta]g_n(x)(1, r(x))^T(1, r(x))\mathrm{d}x \xrightarrow{P} \int h_\theta(x)(1, r(x))^T(1, r(x))\mathrm{d}x. \tag{A.7}$$

As a result, (A.6) reduces to (A.2).

(ii) Suppose (D6) holds. Then by (3.2),

$$
\begin{aligned}
E&\left|\int |r_i(x)r_j(x)r_k(x)| \sup_{t\in\Theta\cap B(\theta,\epsilon)} \exp[(1,r(x))t]g_n(x)\mathrm{d}x\right| = \int |r_i(x)r_j(x)r_k(x)| \sup_{t\in\Theta\cap B(\theta,\epsilon)} \exp[(1,r(x))t]E[g_n(x)]\mathrm{d}x\\
&= \int |r_i(x)r_j(x)r_k(x)| \sup_{t\in\Theta\cap B(\theta,\epsilon)} \exp[(1,r(x))t]\int_{-a_0}^{a_0} K_0(t)g(x+tb_n)\mathrm{d}t\mathrm{d}x\\
&\le \int |r_i(x)r_j(x)r_k(x)| \sup_{t\in\Theta\cap B(\theta,\epsilon)} \exp[(1,r(x))t]\sup_{|t|\le a_0} g(x+tb_n)\mathrm{d}x\\
&= O(1).
\end{aligned}
\tag{A.8}
$$

Therefore, $\int |r_i(x)r_j(x)r_k(x)| \sup_{t\in\Theta\cap B(\theta,\epsilon)} \exp[(1,r(x)\theta_t)]g_n(x)\mathrm{d}x = O_P(1)$ and thus (A.5) holds. Similarly,

$$
\begin{aligned}
E&\left[\int |r_i(x)r_j(x)r_k(x)| \sup_{t\in\Theta\cap B(\theta,\epsilon)} \exp\left[\frac{1}{2}(1,r(x))t\right]g_n^{1/2}(x)h_m^{1/2}(x)\mathrm{d}x\right]^2\\
&\le E\left[\int |r_i(x)r_j(x)r_k(x)|^2 \sup_{t\in\Theta\cap B(\theta,\epsilon)} \exp[(1,r(x))t]g_n(x)\mathrm{d}x\int h_m(x)\mathrm{d}x\right]\\
&= \int |r_i(x)r_j(x)r_k(x)|^2 \sup_{t\in\Theta\cap B(\theta,\epsilon)} \exp[(1,r(x))t]E[g_n(x)]\mathrm{d}x\\
&\le \int |r_i(x)r_j(x)r_k(x)|^2 \sup_{t\in\Theta\cap B(\theta,\epsilon)} \exp[(1,r(x))t]\sup_{|t|\le a_0} g(x+tb_n)\mathrm{d}x\\
&= O(1)
\end{aligned}
$$

and hence (A.3) holds. As a result (A.5) holds. By (3.1), (3.2) and a similar argument as in (A.8),

$$
\begin{aligned}
&\left|\int r_i(x)r_j(x)\exp\left[\frac{1}{2}(1,r(x))\theta\right]\left\{g_n^{1/2}(x)h_m^{1/2}(x)-g^{1/2}(x)h_\theta^{1/2}(x)\right\}\mathrm{d}x\right|\\
&\le \int |r_i(x)r_j(x)|\exp\left[\frac{1}{2}(1,r(x))\theta\right]g_n^{1/2}(x)|h_m^{1/2}(x)-h_\theta^{1/2}(x)|\mathrm{d}x\\
&\quad+\int |r_i(x)r_j(x)|\exp\left[\frac{1}{2}(1,r(x))\theta\right]h_\theta^{1/2}(x)|g_n^{1/2}(x)-g^{1/2}(x)|\mathrm{d}x\\
&\le \left[\int |r_i(x)r_j(x)|^2\exp[(1,r(x))\theta]g_n(x)\mathrm{d}x\right]^{1/2}\left[\int \left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)^2\mathrm{d}x\right]^{1/2}\\
&\quad+\left[\int |r_i(x)r_j(x)|^2\exp[(1,r(x))\theta]h_\theta(x)\mathrm{d}x\right]^{1/2}\left[\int \left(g_n^{1/2}(x)-g^{1/2}(x)\right)^2\mathrm{d}x\right]^{1/2}\\
&= O_P\left(\left[\int\left(h_m^{1/2}(x)-h_\theta^{1/2}(x)\right)^2\mathrm{d}x\right]^{1/2}\right)+O\left(\left[\int\left(g_n^{1/2}(x)-g^{1/2}(x)\right)^2\mathrm{d}x\right]^{1/2}\right)
\end{aligned}
$$

and thus (A.6) holds. By (3.1), (3.3) and using a similar argument as in (A.8),

$$
\begin{aligned}
&\left|\int r_i(x)r_j(x)\exp[(1,r(x))\theta]\big(g_n(x)-g(x)\big)\mathrm{d}x\right|^2\\
&\le \left[\int |r_i(x)r_j(x)|\exp[(1,r(x))\theta]\left|(g_n^{1/2}(x)-g^{1/2}(x))(g_n^{1/2}(x)+g^{1/2}(x))\right|\mathrm{d}x\right]^2\\
&\le \int |r_i(x)r_j(x)|^2\exp[2(1,r(x))\theta](g_n^{1/2}(x)+g^{1/2}(x))^2\mathrm{d}x\int (g_n^{1/2}(x)-g^{1/2}(x))^2\mathrm{d}x\\
&\le 2\left[\int |r_i(x)r_j(x)|^2\exp[2(1,r(x))\theta]g_n(x)+\int |r_i(x)r_j(x)|^2\exp[(1,r(x))\theta]h_\theta(x)\mathrm{d}x\right]\cdot\int (g_n^{1/2}(x)-g^{1/2}(x))^2\mathrm{d}x\\
&= O\left(\int (g_n^{1/2}(x)-g^{1/2}(x))^2\mathrm{d}x\right)\\
&\xrightarrow{P} 0,
\end{aligned}
$$

i.e. (A.7) holds. As a result, (A.5) reduces to (A.2). □

References

[1] P.J. Bickel, C. Klaassen, J. Wellner, Efficient and Adaptive Estimation for Semiparametric Models, Johns Hopkins University Press, Baltimore, 1993.
[2] A.W. Van der Vaart, Asymptotic Statistics, Cambridge University Press, 1998.
[3] P.J. Bickel, J. Kwon, Inference for semiparametric models: Some questions and an answer, Statist. Sinica 11 (2001) 863–960.
[4] J. Forrester, W. Hooper, H. Peng, A. Schick, On the construction of efficient estimators in semiparametric models, Statist. Decisions 21 (2003) 109–138.
[5] R.L. Prentice, R. Pyke, Logistic disease incidence models and case-control studies, Biometrika 66 (1979) 403–411.
[6] J.A. Anderson, Separate sample logistic discrimination, Biometrika 59 (1972) 19–35.
[7] J.A. Anderson, Multivariate logistic compounds, Biometrika 66 (1979) 17–26.
[8] N. Breslow, N.E. Day, Statistical Methods in Cancer Research, in: The Analysis of Case-Control Studies, vol. 1, IARC, Lyon, 1980.
[9] J. Qin, Empirical likelihood in biased sample problems, Ann. Statist. 21 (1993) 1182–1196.
[10] J. Qin, B. Zhang, A goodness of fit test for logistic regression models based on case-control data, Biometrika 84 (1997) 609–618.
[11] B. Zhang, Quantile estimation under a two-sample semi-parametric model, Bernoulli 6 (2000) 491–511.
[12] R. Beran, Minimum Hellinger distance estimators for parametric models, Ann. Statist. 5 (1977) 445–463.
[13] D.L. Donoho, R.C. Liu, The "automatic" robustness of minimum distance functionals, Ann. Statist. 16 (1988) 552–586.
[14] B.G. Lindsay, Efficiency versus robustness: The case for minimum Hellinger distance and related methods, Ann. Statist. 22 (1994) 1081–1114.
[15] C.R. Stather, Robust statistical inference using hellinger distance methods, Ph.D. dissertation, LaTrobe University, Australia, 1981.
[16] R.N. Tamura, D.D. Boos, Minimum Hellinger distance estimation for multivariate location and covariance, J. Amer. Statist. Assoc. 81 (1986) 223–229.
[17] D.G. Simpson, Minimum Hellinger distance estimation for the analysis of count data, J. Amer. Statist. Assoc. 82 (1987) 802–807.
[18] S. Yang, Minimum Hellinger distance estimation of parameter in the random censorship model, Ann. Statist. 19 (1991) 579–602.
[19] Z. Ying, Minimum Hellinger-type distance estimation for censored data, Ann. Statist. 20 (1992) 1361–1390.
[20] W.A. Woodward, P. Whitney, P.W. Eslinger, Minimum Hellinger distance estimation of mixture proportions, J. Statist. Plann. Inference 48 (1995) 303–319.
[21] A. Cutler, O.I. Cordero-Braña, Minimum Hellinger distance estimation for finite mixture models, J. Amer. Statist. Assoc. 91 (1996) 1716–1723.
[22] D. Karlis, E. Xekalaki, Robust inference for finite Poisson mixtures, J. Statist. Plann. Inference 93 (2001) 93–115.
[23] T.N. Sriram, A.N. Vidyashankar, Minimum Hellinger distance estimation for supercritical Galton–Watson processes, Statist. Probab. Lett. 50 (2000) 331–342.
[24] Mi-Ja Woo, T.N. Sriram, Robust estimation of mixture complexity, J. Amer. Statist. Assoc. 101 (2006) 1475–1486.
[25] Mi-Ja Woo, T.N. Sriram, Robust estimation of mixture complexity for count data, Comput. Statist. Data Anal. 51 (2007) 4379–4392.
[26] Z. Lu, Y.V. Hui, A.H. Lee, Minimum Hellinger distance estimation for finite mixtures of Poisson regression models and its applications, Biometrics 59 (2003) 1016–1026.
[27] R.J. Karunamuni, J. Wu, Minimum Hellinger distance estimation in a nonparametric mixture model, J. Statist. Plann. Inference 139 (2009) 1118–1133.
[28] J. Wu, R.J. Karunamuni, On minimum Hellinger distance estimation, Canadian J. Statist. 37 (2009) 514–533.
[29] J. Wu, Minimum Hellinger distance estimation in semiparametric models, Ph.D. Dissertation, University of Alberta, 2007.
[30] P.J. Rousseeuw, C. Croux, Alternatives to the median absolute deviation, J. Amer. Statist. Assoc. 88 (1993) 1273–1283.
[31] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, Wiley, New York, 1989.